ESD-TDR-64-530

MATHEMATICAL PROGRAMMING TECHNIQUES FOR INFORMATION SYSTEM DESIGN

TECHNICAL DOCUMENTARY REPORT NO. ESD-TDR-64-530
July 1964

U. O. Gagliardi
C. G. Ying
L. G. Holt

DIRECTORATE OF COMPUTERS
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

Project 2804

(Prepared under contract AF19(628)2830 by Dunlap & Associates, Inc., Darien, Conn.)

FOREWORD

The authors wish to acknowledge the valuable assistance they received from Lt. John B. Curtis, USAF, who has been instrumental in monitoring meetings with ESD and Mitre personnel and in supplying information and suggestions which helped the project.

ABSTRACT

The problem of establishing the effectiveness of an information system is considered. An effectiveness measure suggested by recent development in statistical decision theory is presented. Sample evaluations or system designs are used to illustrate here the adoption of such a measure which allows selecting the parameters of the system in a manner consistent with the user preference.

This Technical Documentary Report has been reviewed and is approved.

John B. Curtis
JOHN B. CURTIS
2d Lt, USAF
Contract Monitor

TABLE OF CONTENTS

KEY WORDS LIST

Command Control Systems

Statistical Analysis

Computers

Operations Research

Communication Systems

1. THE RATIONALE FOR ADOPTING THE VIEWPOINT OF STATISTICAL DECISION THEORY IN EVALUATING INFORMATION SYSTEMS

1.1 The Design Process

A system design in its final form is the result of an aggregate of design decisions concerning the specifications of its components and the structure of their connections.

The information concerning the merits of available design alternatives is obtained either through logical implications of a performance model for the family of systems in question or through experimental or historical evidence which indicates which system configurations have been successful in the past. A case in which design relied quite heavily on the experimental approach is represented by the design of modern automobiles. The determinations of increasingly satisfactory power ratings, number and geometry of the piston assembly, stroke and bore dimensions, compression ratio, type of cycle, etc., all have resulted from a long history of testing alternative designs.

In the case of the engineering of information systems, reliance on such an experimental weeding of unsatisfactory system's designs is not altogether desirable for the following reasons:

- Information systems tend to be strongly differentiated, sometimes to the point of being one of a kind.
- Information systems testing can be very expensive or, in the military, sometimes not feasible.
- The cost resulting from design errors can be very high.
- Information systems become obsolete too rapidly for the trial and error approach.

If in "vivo" testing by experiments is to be avoided, then one is left with the alternative of acquiring the knowledge concerning various designs by modeling. Briefly, a model of a system characterizes, in the light of a given set of design assumptions, the transformation of the inputs of the system into its outputs and thereby arrives at the computation of a performance measure for the system.

1.2 The Value of Information

Information in general denotes a set of potential messages associated with a given channel or system of information. It can be viewed as something which informs us about the state of a certain environment so that the uncertainty associated with such an environment can be expected to be reduced if not completely eliminated. Our desire for the reduction of uncertainty concerning the environment can be based on either purely intellectual motivations (e.g., to satisfy the curiosity of the mind) or the need for making decisions with the hope of achieving some economic goals. In situations where the economic motivation is a dominating factor, the value of an information system has to be assessed by figuring out how much we expect it to help us to attain our goals. Since information is the output of an information system, in order to determine the value associated with an information system, we have to know the nature of our decision task and how the provided information is utilized in the decision process.

1.3 The Inadequacy of Applying the Traditional Information Theory in Evaluating Information Systems

Let Z be a random variable that assumes N values, denoted by $z_1 \ldots z_N$, respectively; and let $P(z_i)$ be the probability of z_i. Then a statistical parameter called the "entropy" associated with the random variable Z, denoted by H(Z), is defined as

$$H(Z) = -\sum_i P(z_i) \cdot \log P(z_i)$$

When the z's represent the various states of nature, the entropy measures, in some sense, the degree of uncertainty associated with the nature. When the z's represent the set of potential messages associated with an information channel, it is used by Shannon* as a measure of "the amount of information."

Let $X = (x_1 \ldots x_N)$ denote the set of N possible states of nature, and $Y = (y_1 \ldots y_N)$ denote the set of potential messages associated with an information channel. Let $H(X|y_j) = -\sum_i P(x_i|y_j) \cdot \log P(x_i|y_j)$, where $P(x_i|y_j)$ is the conditional probability of x_i given y_j. $H(X|y_j)$ can be

---

* Shannon, C. "A Mathematical Theory of Communication," Bell System Technical Journal, 1948.

interpreted as a measure of the amount of uncertainty (about the nature) remaining after receiving the j-th message. Then the equivocation of the channel Y with respect to the source X, denoted by $H(X|Y)$, is defined as

$$H(X|Y) = \sum_j P(y_j) \cdot H(X|y_j)$$

The transmission rate of Y with respect to X, denoted by $R(Y, X)$, is defined as

$$R(Y, X) = H(X) - H(X|Y)$$

and the channel capacity, denoted by $C(Y)$, is defined as

$$C(Y) = \underset{X}{\text{Max}}\ R(Y, X)$$

Since the derivation of the channel capacity (and that of transmission rate, etc.) does not take into consideration the user's utility structure anywhere, it is clear that it cannot adequately represent the value of an information channel to a particular user faced with a particular decision task -- although under certain special assumptions concerning the user's utility structure, the channel capacity can be made to correspond to the value of an information channel to the user.*

What then does the entropy associated with an information channel represent? According to Marschak,** since the entropy usually increases with the number of distinct potential messages, and the larger the number of distinct potential messages, the larger the number of symbols needed at a minimum to distinguish the messages, the entropy represents, more appropriately, the cost of constructing an information instrument. This perhaps explains, at least partially, the fact that the entropy concept was first proposed by the people at Bell Lab. A producer of information instruments cannot hope to take into consideration the various needs of different users of its product; however, he does concern himself about it costs associated with producing various information instruments.

---

* See Kelly, J. "A New Interpretation of Information Rate," Bell System Technical Journal, 1956.

** Marschak, J. "Remarks on the Economics of Information," Contributions to Scientific Research in Management, 1960.

1.4 Statistical Decision Theory

Statistical decision theory is concerned with the derivation of an optimal decision rule (in the face of uncertainty) based on

- The decision maker's utility structure.
- The decision maker's a priori probability distribution over the states of nature.
- An experiment which generates observations.
- A rule of revising the prior distribution upon receiving an observation.
- A definition of what constitutes an optimal decision rule.

In deriving such an optimal decision rule, one introduces a measure of performance over the set of all possible decision rules which is the expected value of some suitable function of the utilities. A natural application of this theory in evaluating information systems is to model the information system as a scattering process (an experiment) in which any particular state of nature can give rise to several possible messages (observations), and the value associated with an information system is obtained by computing the expected gain (using the optimal decision rule) as a result of employing the system.

It then follows that an appropriate measure of effectiveness of a given information system to a particular user is the net gain in the expected utility, resulting from employing the information system as an aid in decision-making processes over and above the expected utility which results when no information system is employed.

If the effectiveness of an information system is defined in the manner outlined above, it will be logically determined by, and thus consistent with, the utilities chosen and the statistical hypotheses concerning the prior uncertainty about the environment.

Since effectiveness-cost equilibrium will determine the design recommendations, we ultimately have in this theory a means to make design recommendations which are consistent with the aforementioned elements. On the other hand, if one would adopt, following a common

practice, measures of performance of components as the basis for performing trade-off analyses, one may recommend a design which does not agree with the user preference. Such design may in fact imply a utility structure which the user would not agree with, if rendered explicit to him. Worse yet, the design so obtained may imply mutually contradictory statements of preference so that there is no utility structure which is consistent with it.

2. THE CHOICE OF PREMISES

The method developed in these notes allows us to determine the value associated with a set of design specifications for an information system which is consistent with the following premises:

- The utility associated with an act-state pair (A^k, X^i) is the real number $u(A^k, X^i)$, where A^k ranges over all the acts open to the decision maker, and X^i ranges over all the possible states of the environment.

- The prior probability that the environment will be in state X^i is the probability measure $P(X^i)$.

- A definition of what constitutes an optimal decision rule.

Some brief comments on the selection of particular instances of these premises are in order. The search for optimal decision rules necessarily entails the establishment of a criterion of optimality. A criterion of optimality can be established by specifying a particular objective function, in our case, a particular utility structure.

An objective function, in general, and a utility structure, in particular, can be specified in the following ways:

- It is specified by a ranking scheme among all the possible outcomes based on the preference of the decision maker involved.

- It is specified by a ranking scheme based on a so-called "average preference."

- It is specified arbitrarily.

The last possibility can be ignored because of its attendant risks of irrelevancy and inconsistency. The second possibility can be eliminated because it is difficult, in general, to exhibit an "average preference" which would obtain the consensus of most decision makers and because the experimental survey of opinion is likely to be expensive or even infeasible. This leaves us only the first possibility. Admittedly, the ranking scheme then involves a subjective value judgment of a person

or of a nonrepresentative group* and thus may lack universality and may be highly variable with time.

The lack of universality is immaterial since one is designing a system for a given user who may not agree, or who even may have to disagree, with some other user due to his different goal systems. For example, the information delivered to a wing commander may have to be evaluated in a way which is substantially different from that used for the evaluation of the information delivered to a squadron commander; the latter's goal of destroying individual enemy fighters may be only a subgoal of very minor importance for the former. Actually, the point can be made that one of the uses of our method is to verify whether the utility structures of two distinct users are sufficiently congruent at the design recommendation level to warrant a joint design and sharing of costs, or whether it is better to keep two separate systems because the cost of compromising, in terms of lowered performance, is too high.

The instability, in time, of the value structure of the user can be reduced by proper selection of the decision-making context** for the evaluation, but it cannot be eliminated completely because of factors such as developments in weapon, sensing, communication, etc., technologies. In any event, if the value judgment is not made at the level of specifying the utility structure, it will be made at a lower level, that of specifying the system's technical characteristics and, as said before, such value judgments are equivalent to value judgments made about utilities with the added disadvantage that the various system parameter's judgments may not be mutually consistent. Since the method translates the value judgments of the user concerning the value to him of certain actions in the face of certain environmental situations into recommendations concerning the technical parameters of the system, it has the added advantage of letting the user make the value judgment.

What is meant here is that the military commander is more familiar with his world and act variables than with a context made up of error rates, channel capacities, rate of information transfer, storage capacities, etc. He has, then, a capability of making an informed judgment within the decision-making context, but may have no real feel for what it is worth to him to have one rather than another set of specifications for the information system, especially if it is an involved system. This in turn may result in agreeing to design recommendations, which, as repeatedly pointed out, may not really agree with the decision maker's preference.

---

* A decision maker can be either an individual or a group.

**Toda, M. and Shuford, E. H., Jr. TDR No. ESD-TDR-63-622, Oct. 1963

Since the proposed method relies heavily on automating the computation of the system effectiveness, it would permit us to quickly obtain a view of the effectiveness of many alternative system configurations in the light of a single utility and prior probability premise. This would in turn allow comparison of different alternative premises to test the stability of the design recommendations with respect to changes of the premises. In other words, one could obtain indications of the more critical elements in the premises. A better informed judgment on the part of the user could then be obtained for those elements of either the utility structure or the prior distribution, which prove to be most critical.

The task at hand can be considerably reduced if one considers that:

- The utility structure is, in general, an array of induced utilities* which depend on fewer absolute utilities and a set of probabilities of obtaining a goal given that a given subgoal has been achieved. One needs, then, to test the design-criticality of the absolute utilities. (The success probabilities are, in general, known.)

- The prior probabilities may be expressed in terms of a limited number of parameters or there may be sufficient statistical information to determine them.

Finally, with regard to the selection of a definition of an optimal decision rule, we can say that in this report we will adopt a Bayesian viewpoint, i.e., a decision rule is said to be optimal if its adoption maximizes the expected utility, using as weights the posterior probabilities which are computed by employing Bayes theorem.

*Toda, M. and Shuford, E.H., Jr. TDR No. ESD-TDR-63-622, Oct 1963.

3. THE BAYESIAN EFFECTIVENESS OF INFORMATION SYSTEMS

3.1 The Bayesian Value of an Information System

The usual model of decision making under uncertainty assumes that there are certain states of nature that are relevant to our decision, certain acts that are open to us for choice, and a utility index associated with each act-state pair.

Let X^j denote the j-th state of nature, $j = 1, \ldots, N$; A^i denote the i-th act open to us, $i = 1, \ldots, L$; and u_{ij} be the utility index assigned to the act-state pair (A^i, X^j), $u_{ij} = U(A^i, X^j)$.

An information structure can be most conveniently characterized as follows:

$$\begin{array}{c} \\ X^1 \\ \cdot \\ \cdot \\ \cdot \\ X^N \end{array} \begin{array}{c} \begin{array}{ccccc} Y^1 & \cdot & \cdot & \cdot & Y^M \end{array} \\ \begin{bmatrix} q_{11} & \cdot & \cdot & \cdot & q_{1M} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ q_{N1} & \cdot & \cdot & \cdot & q_{NM} \end{bmatrix} \end{array}$$

where Y^k, $k = 1, \ldots, M$, is the k-th message transmitted to us by the information system, and $q_{ik} = P(Y^k | X^i)$ is the conditional probability of the k-th message given the fact that the true state of the nature is X^i.

A rule which assigns an act to each of the possible messages is called a decision rule. We shall denote it by $A = \alpha(Y)$.

The Bayesian decision rule assumes the following things: (1) There is a certain prior probability associated with each state of nature; we shall denote it by $P(X^j)$, $j = 1, \ldots, N$. (2) For each message observed, an a posteriori probability distribution over the states of nature can be derived by using the Bayes theorem. Let $P(X^j | Y^k)$ denote the posterior probability of X^j given the fact that Y^k has been observed. Then,

$$P(X^j | Y^k) = P(X^j) \cdot P(Y^k | X^j) \Big/ \sum_i P(X^i) \cdot P(Y^k | X^i).$$

(3) Let $V(A^i|Y^k) = \sum_j P(X^j|Y^k)\, u_{ij}$ be the expected value of A^i given the fact that Y^k has been observed. Then the Bayesian decision rule says that for each message Y^k one should select the act $A = \hat{\alpha}(Y^k)$ such that

$$V[\hat{\alpha}(Y^k)|Y^k] = \max_A V(A|Y^k).$$

Let $P(Y^i) = \sum_j P(X^j) \cdot P(Y^i|X^j)$ be the probability of observing the i-th message given the prior probability distribution over X and the information system $\varkappa$. Then the Bayesian value of $\varkappa$ is

$$\hat{V}(\varkappa) = \sum_{i=1}^{M} P(Y^i) \sum_{j=1}^{N} P(X^j|Y^i)\, u\,[\hat{\alpha}(Y^i), X^j]. \qquad (3.1\text{-}1)$$

3.2 Some Examples of Information Systems

3.2.1 Perfect Information System

$$\begin{array}{c} \\ X^1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ X^N \end{array}
\begin{array}{c} \begin{array}{ccccc} Y^1 & \cdot & \cdot & \cdot & Y^M \end{array} \\
\begin{bmatrix} 1 & 0 & \cdot & \cdot & 0 \\ 0 & 1 & \cdot & \cdot & 0 \\ \cdot & & \cdot & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & \cdot & \cdot \\ 0 & \cdot & & 0 & 1 \end{bmatrix} \end{array}$$

$$q_{ij} = 1 \quad \text{if} \quad i = j$$

$$q_{ij} = 0 \quad \text{if} \quad i \neq j$$

The perfect information system is interesting because it serves as a least upper bound for the Bayesian value of an information system, which is a measure of the best (according to the Bayesian criterion) a decision maker can do with the aid of a perfect information system. Such a bound will tell us what the maximum gain is that can be expected by improving the precision of an information system and whether an effort to improve the information justifies its cost.

3.2.2 Null Information System

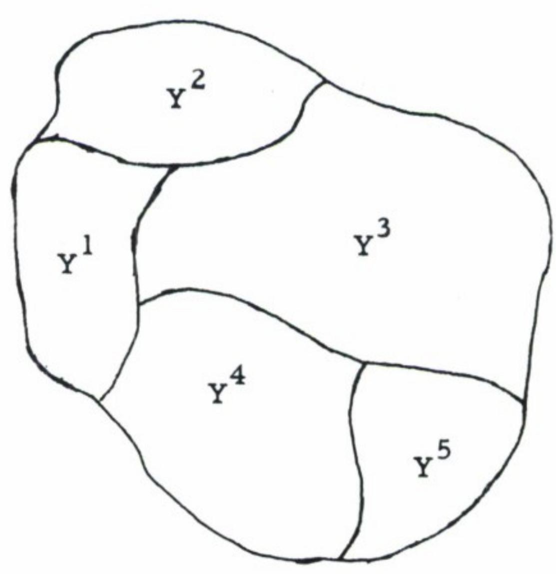
$$
\begin{array}{cc}
 & \begin{array}{ccccc} Y^1 & \cdot & \cdot & \cdot & Y^M \end{array} \\
\begin{array}{c} X^1 \\ \cdot \\ \cdot \\ \cdot \\ X^N \end{array} &
\begin{bmatrix}
q_{11} & \cdot & \cdot & \cdot & q_{1M} \\
\cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot \\
q_{N1} & \cdot & \cdot & \cdot & q_{NM}
\end{bmatrix}
\end{array}
$$

$$q_{ij} = q_{kj} \quad \text{for} \quad j = 1, \ldots, M.$$

The null information system is interesting because it serves as a greatest lower bound for the Bayesian value of an information system, which is a measure of the best (according to the Bayesian criterion) a decision maker can do without the aid of an information system.

3.2.3 Imperfect Information System as a Partition of X

This is a situation where the information system serves as a scheme to partition the set of all states of nature, which is usually the result of coding the states of nature using fewer distinct messages than the total number of distinct states of nature.



$$
\begin{array}{cc}
 & \begin{array}{ccccccc} Y^1 & Y^2 & Y^3 & \cdot & \cdot & \cdot & Y^M \end{array} \\
\begin{array}{c} X^1 \\ \vdots \\ X^{N_1} \\ X^{N_1+1} \\ \vdots \\ X^{N_2} \\ X^{N_2+1} \\ \vdots \\ X^N \end{array} &
\begin{bmatrix}
1 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\
0 & 1 & \cdot & \cdot & \cdot & \cdot & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\cdot & 1 & 0 & \cdot & \cdot & \cdot & 0 \\
\cdot & 0 & 1 & \cdot & \cdot & \cdot & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdot & \cdot & \cdot & 1
\end{bmatrix}
\end{array}
$$

$q_{ij} = 1$ if $N_{j-1} < i \leqq N_j$

$q_{ij} = 0$ otherwise

where $N_0 = 0$

3.3 The Bayesian Effectiveness

As defined in Section 3.1, the Bayesian decision rule selects the act $A = \hat{\alpha}(Y^k)$ such that

$$V[\hat{\alpha}(Y^k) \mid Y^k] = \max_A V(A \mid Y^k)$$

i.e., such that

$$\sum_{j=1}^{N} P(X^j \mid Y^k)\, u\,[\hat{\alpha}(Y^k), X^j] = \max_A \left[\sum_{j=1}^{N} P(X^j \mid Y^k)\, u\,(A, X^j) \right]$$

Then the Bayesian value of the information system χ is:

$$\hat{V}(\chi) = \sum_{i=1}^{M} P(Y^i) \max_A \left[\sum_{j=1}^{N} P(X^j \mid Y^i)\, u\,(A, X^j) \right]$$

By Bayes theorem we have that

$$P(X^j \mid Y^i) = \frac{P(Y^i \mid X^j)\, P(X^j)}{P(Y^i)}$$

Thus

$$\hat{V}(\chi) = \sum_{i=1}^{M} P(Y^i) \max_A \left[\sum_{j=1}^{N} \frac{P(Y^i \mid X^j)\, P(X^j)}{P(Y^i)}\; u\,(A, X^j) \right] =$$

$$\sum_{i=1}^{M} \max_A \left[\sum_{j=1}^{N} P(Y^i \mid X^j)\, \bar{u}\,(A, X^j) \right] \qquad (3.3\text{-}1)$$

where the $\{P(Y^i \mid X^j)\}$ can be directly specified once the statistical nature of the error processes in the information system is known, and the functions $\bar{u}(A, X^j)$ are given by

$$\bar{u}(A, X^j) = u(A, X^j)\, P(X^j)$$

This suggests that the effectiveness of an information system, $E(\chi)$, may be considered as the net gain in the expected utility, resulting from employing the information system as an aid in decision-making processes over and above the expected utility which results when no information system is employed. That is:

$$E(\chi) = \hat{V}(\chi) - \hat{V}(\chi_o) \qquad (3.3\text{-}2)$$

where χ_o is the null information system defined by

$$P(X^j | Y^i) = P(X^j)$$

for all i and j.

The value of $\hat{V}(\chi_o)$ is given by:

$$\hat{V}(\chi_o) = \sum_{i=1}^{M} P(Y^i) \max_{A} \left[\sum_{j=1}^{N} P(X^j)\, u(A, X^j) \right] =$$

$$\max_{A} \left[\sum_{j=1}^{N} P(X^j)\, u(A, X^j) \right] =$$

$$\max_{A} \left[\sum_{j=1}^{N} \bar{u}(A, X^j) \right] \qquad (3.3\text{-}3)$$

Equations (3.3-1), (3.3-2), and (3.3-3) then allow the computation of the Bayesian effectiveness of the information system χ characterized by the "noise" probabilities $\{P(Y^i|X^j)\}$ in the light of the utility functions $\{u(A, X^j)\}$ and prior distribution $\{P(X^j)\}$.

The solution of equations (3.3-1) and (3.3-3) constitute problems of the mathematical programming type. In fact, in both cases one is maximizing a convex combination of the functions $u(A, X^j)$. Such a convex combination (for a given Y^i in the case of equation (3.3-1)) is a function of A. A in turn can be limited to a suitably defined region of an n-dimensional vector space. In Appendix A we present a special case of the damage assessment function which results in a nonlinear program for which the Kuhn-Tucker conditions generate a resolvent algorithm. Whenever it is possible to supply an efficient resolvent algorithm for equations (3.3-1) and (3.3-3), one is then in the position to evaluate the effectiveness of the information system using equation (3.3-2).

Unfortunately, it is frequently the case that either the act space or the utility functions are such that no efficient algorithmic solution exists besides the exhaustion of alternatives.

In the following section we shall deal with the case in which the continuous functions of a $\{u(A, X^j)\}$ are replaced by the columns of a utility matrix $\{u(A^k, X^j)\}$. Such discrete formulation is useful for the programming of a digital computer to evaluate all the alternatives, which in most practical cases will be the only method available.

3. 3. 1 The Bayesian Effectiveness (Discrete Formulation)

In subsequent sections, the following notations will be used:

$$U = \begin{bmatrix} u_{11} & \cdot & \cdot & \cdot & u_{1N} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ u_{L1} & \cdot & \cdot & \cdot & u_{LN} \end{bmatrix}$$

where $u_{ki} = U(A^k, X^i)$

$$Q = \begin{bmatrix} q_{11} & \cdot & \cdot & \cdot & q_{1M} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ q_{N1} & \cdot & \cdot & \cdot & q_{NM} \end{bmatrix}$$

where $q_{ij} = P(Y^j | X^i)$

$$P = \begin{bmatrix} p_{11} & \cdot & \cdot & \cdot & p_{1M} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ p_{N1} & \cdot & \cdot & \cdot & p_{NM} \end{bmatrix}$$

where $p_{ij} = P(X^i|Y^j)$

$$= \frac{P(X^i) \cdot P(Y^j|X^i)}{\sum_{k=1}^{N} P(X^k) \cdot P(Y^j|X^k)}$$

$$= \frac{P(X^i) \cdot P(Y^j|X^i)}{P(Y^j)}$$

The j-th column of P, denoted by $[P]_j$,

$$\begin{bmatrix} p_{1j} \\ \cdot \\ \cdot \\ \cdot \\ p_{Nj} \end{bmatrix} = \begin{bmatrix} P(X^i|Y^j) \\ \cdot \\ \cdot \\ \cdot \\ P(X^N|Y^j) \end{bmatrix}$$

is the conditional probability distribution over the states of nature if Y^j has been observed. It is then clear that the j-th column of UP, denoted by $[UP]_j$, is the set of expected utilities associated with various acts conditional on the occurrence of Y^j.

We shall now define the operator *. Let [A] be a column vector.

$$\begin{bmatrix} a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_M \end{bmatrix} . \qquad \text{Then } [A]^* = \max_i \{a_i\}$$

Let A be a matrix,

$$\begin{bmatrix} a_{11} & \cdot & \cdot & \cdot & a_{1N} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ a_{M1} & \cdot & \cdot & \cdot & a_{MN} \end{bmatrix}$$

Then $A^* = ([A]^*_1 \ . \ . \ . \ [A]^*_N)$, where $[A]_j$ denotes the j-th column of A.

With the aid of the operator *, we can define the Bayesian decision rule as $\hat{\alpha}(Y^j) = A^{\hat{i}}$ such that $V(A^{\hat{i}} | Y^j) = [UP]^*_j$. Then the Bayesian value of an information system $\varkappa$ is given by

$$\hat{V}(\varkappa) = \sum_{j=1}^{M} P(Y^j)\,[UP]_j^*$$

where $P(Y^j) = \sum_{k=1}^{N} P(X^k) \cdot P(Y^j|X^k)$.

Let $\hat{E}(\varkappa)$ be the Bayesian effectiveness associated with an information system $\varkappa$. Then

$$\hat{E}(\varkappa) = \hat{V}(\varkappa) - \hat{V}(\varkappa^o)$$

where $\varkappa^o$ denotes the null information system.

Consider the k-th component of $[UP]_j$. It is

$$\sum_{i=1}^{N} U(A^k, X^i)\, P(X^i|Y^j) = \frac{1}{P(Y^j)} \sum_{i=1}^{N} U(A^k, X^i)\, P(X^i)\, P(Y^j|X^i).$$

Let $\overline{U} = UD$

where $D = \begin{bmatrix} P(X^1) & \cdot & \cdot & \cdot & 0 \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & P(X^N) \end{bmatrix}$ is a diagonal matrix whose

diagonal elements are $P(X^1) \ldots P(X^N)$. Then the k-th component of $[UP]_j$ is simply $\frac{1}{P(Y^j)}\,[\overline{U}Q]_{kj}$, where $[\overline{U}Q]_{kj}$ is the kj-th element of $\overline{U}Q$.

$$[UP]_j = \frac{1}{P(Y^j)}\,[\overline{U}Q]_j\,.$$

Since $\left(\frac{1}{P(Y^j)}\,[\overline{U}Q]_j\right)^* = \frac{1}{P(Y^j)}\,[\overline{U}Q]_j^*$,

it follows that

$$V(\varkappa) = \sum_{j=1}^{M} P(Y^j)\ [UP]^*_j$$

$$= \sum_{j=1}^{M} P(Y^j)\ \left(\frac{1}{P(Y^j)}\ [\overline{U}Q]_j\right)^*$$

$$= \sum_{j=1}^{M} [\overline{U}Q]^*_j$$

$$= (\overline{U}Q)^*\ \xi$$

where ξ is a column vector with M components whose values are all equal to 1.

Let P_o be the P matrix associated with the null information system $\varkappa_o$. Since $[UP_o]_j$ is the weighted average--with the weights $\{P(X^i)\}$-- of the columns of U and is independent of j, we shall denote it by $[U_o]$. Then

$$V(\varkappa_o) = \sum_{j=1}^{M} P(Y^j)\ [U_o]^* = [U_o]^*$$

and

$$E(\varkappa) = (\overline{U}Q)^*\ \xi\ -\ [U_o]^* = (\overline{U}Q)^*\ \xi\ -\ (U\pi)^*$$

where π is the vector whose i-th component is the prior probability of X^i, $P(X^i)$.

4. PARTITIONING OF INFORMATION SYSTEMS

Suppose that each state of nature is partitioned into two parts denoted by X and Z, respectively, and each message is partitioned into two parts denoted by Y and W, respectively, such that the following conditions are satisfied:

(1) $P(X, Z) = P(X)\, P(Z)$

(2) $P(Y|X, Z) = P(Y|X, Z')$ for all Z and Z'

(3) $P(W|X, Z) = P(W|X', Z)$ for all X and X'

This represents a situation in which X and Z are statistically independent, and Y and W are the outputs of channels that do not cross talk. Then,

$$P(Y|X) = \sum_Z P(Z)\, P(Y|X, Z)$$

$$= P(Y|X, Z) \sum_Z P(Z)$$

$$= P(Y|X, Z)$$

$$P(W|Z) = \sum_X P(X)\, P(W|X, Z)$$

$$= P(W|X, Z) \sum_X P(X)$$

$$= P(W|X, Z)$$

$$P(W|X) = \sum_Z P(Z)\, P(W|X, Z)$$

$$= \sum_Z P(Z)\, P(W|Z)$$

$$= P(W)$$

$$P(Y|Z) = \sum_X P(X)\, P(Y|X, Z)$$

$$= \sum_X P(X)\, P(Y|X)$$

$$= P(Y)$$

$$P(W|Y) = \sum_{X, Z} P(W|X, Z)\; P(X, Z|Y)$$

$$= \sum_{X, Z} P(W|Z)\, P(X|Y)\, P(Z|Y)$$

$$= \sum_{Z} P(W|Z)\, P(Z|Y) \sum_{X} P(X|Y)$$

$$= \sum_{Z} P(W|Z)\, P(Z|Y)$$

$$= \sum_{Z} P(W|Z)\, P(Z)$$

$$= P(W)$$

for $P(Z|Y) = \dfrac{P(Y|Z)\,P(Z)}{P(Y)} = P(Z)$.

$$P(W, Y) = P(Y)\, P(W|Y)$$

$$= P(Y)\, P(W)$$

So, we see that Y and W are statistically independent, and it follows that

$$P(Y, W|X, Z) = P(Y|X, Z)\, P(W|X, Z)$$

$$= P(Y|X)\, P(W|Z).$$

Therefore, if we order the rows and columns of Q so that states which share a common X^i are adjacent to each other and messages which share a common Y^j are adjacent to each other, then Q can be partitioned as follows:

$$
\begin{array}{c|c}
 & \begin{matrix} Y^1 & \quad\cdots\quad & Y^M \end{matrix} \\
\begin{matrix} X^1 \\ \cdot \\ \cdot \\ \cdot \\ X^N \end{matrix} & \begin{bmatrix} P(Y^1|X^1)R & & P(Y^M|X^1)R \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ P(Y^1|X^N)R & \cdots & P(Y^M|X^N)R \end{bmatrix}
\end{array}
$$

where R is the matrix.

$$
\begin{array}{c|c}
 & \begin{matrix} W^1 & \quad\cdots\quad & W^T \end{matrix} \\
\begin{matrix} Z^1 \\ \cdot \\ \cdot \\ \cdot \\ Z^S \end{matrix} & \begin{bmatrix} P(W^1|Z^1) & & P(W^T|Z^1) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ P(W^1|Z^S) & \cdots & P(W^T|Z^S) \end{bmatrix}
\end{array}
$$

If we partition the matrix $\overline{U}$ by ordering its columns so that states with a common X^i are adjacent to each other, i. e.,

$$
\begin{array}{ccccc}
 & & X^1 & \cdots & X^N \\
\overline{U} & = & [\ \overline{U}^1 & \cdots & \overline{U}^N\ ]
\end{array}
$$

where each of the $\overline{U}^i$ has S columns, then,

$$
\overline{U}Q = [\ \sum_{i=1}^{N} P(Y^1|X^i)\,\overline{U}^i R, \ \ldots, \ \sum_{i=1}^{N} P(Y^M|X^i)\,\overline{U}^i R]
$$

and

$$
(\overline{U}Q)^* = [\ (\sum_{i=1}^{N} P(Y^1|X^i)\,\overline{U}^i R)^*, \ \ldots, (\sum_{i=1}^{N} P(Y^M|X^i)\,\overline{U}^i R)^*\ ]
$$

5. PARTITIONING AND SYSTEM DESIGN

The result obtained in the previous section,

$$(\overline{U}Q)^* = [(\sum_{i=1}^{N} P(Y^1|X^i)\,\overline{U}^i R)^*, \ldots, (\sum_{i=1}^{N} P(Y^M|X^i)\,\overline{U}^i R)^*],$$

can be utilized to shorten the computation of $(\overline{U}Q)^*$ and $(\overline{U}Q)^*\xi$ and thus of the effectiveness of the system when one of the subsystems is considered fixed.

In designing, one may have sufficient reason to consider one subsystem essentially fixed. For example, one may have sufficient grounds to expect that a given subsystem will, in all acceptable designs, be nearly perfect. Then the size of the evaluation effort would be greatly reduced if the subsystem were modeled with an identity matrix with the effectiveness of the over-all system becoming dependent only on the other subsystem design parameters. If, for example, in the above formula we set $P(Y^j|X^i) = \delta_{ij}$ where $\delta_{ij} = 0,\ i \neq j;\ \delta_{ij} = 1,\ i = j,\ M = N$, i.e., if the (X, Y) subsystem is assumed to be perfect, then

$$(\overline{U}Q)^* = [(\overline{U}^1 R)^* \ldots . (\overline{U}^N R)^*]$$

An alternative formula can be obtained by identifying the perfect subsystem with the (Z, W) subsystem, in which case $R = I$, and

$$(\overline{U}Q)^* = [(\sum_{i=1}^{N} P(Y^1|X^i)\,\overline{U}^i)^* \ldots . (\sum_{i=1}^{N} P(Y^M|X^i)\,\overline{U}^i)^*]$$

Partitioning the system into the two subsystems (X, Y) with model P and (Z, W) with model R is a convenient device to explore the behavior of system effectiveness when only one of the subsystems is varied. In particular, if one is interested in determining if one can afford to ignore certain environmental variables, one must determine what happens to the effectiveness of the over-all system as one of the two subsystems is reduced to a null system, i.e., the over-all system is reduced to the remaining subsystem.

Suppose that the system to be eliminated is the (X, Y) subsystem. Then for it we have that,

$$P(Y^1|X^i) = P(Y^1) \qquad \text{for all i's}$$

$$P(Y^2|X^i) = P(Y^2) \qquad \text{for all i's}$$

$$\vdots$$

$$P(Y^M|X^i) = P(Y^M) \qquad \text{for all i's}$$

Then,

$$(\overline{U}Q)^* = [(\sum_{i=1}^{N} P(Y^1)\,\overline{U}^i R)^*, \ldots, (\sum_{i=1}^{N} P(Y^M)\,\overline{U}^i R)^*]$$

$$= [P(Y^1)\,\{(\sum_{i=1}^{N} \overline{U}^i) R\}^*, \ldots, P(Y^M)\{(\sum_{i=1}^{N} \overline{U}^i) R\}^*]$$

In computing the effectiveness of Q, we have to compute $(\overline{U}Q)^*_\xi$, i. e., the sum of the components of the row vector $(\overline{U}Q)^*$, but this sum is equal to the sum of the components of the vector,

$$P(Y^1)\,[(\sum_{i=1}^{N} \overline{U}^i) R]^* + P(Y^2)\,[(\sum_{i=1}^{N} \overline{U}^i) R]^* + \ldots\ldots$$

$$+ P(Y^M)\,[(\sum_{i=1}^{N} \overline{U}^i) R]^* = [(\sum_{i=1}^{N} \overline{U}^i) R]^* = (\overline{U}^\dagger R)^*$$

i. e., the effectiveness of subsystem (Z, W) operating alone can be computed in the usual manner by constructing a new $\overline{U}$ matrix by summing together all the columns of $\overline{U}$ which correspond to a given Z^P. This of course is an expected result.

If $(\overline{U}Q)^*_{P_{\text{perfect}}} = [(\overline{U}^1 R)^*_, (\overline{U}^2 R)^* \ldots (\overline{U}^N R)^*]$ and $(\overline{U}Q)^*_{P_{\text{null}}} = (\overline{U}^\dagger R)^*$ are used, one can compute an upper bound for the improvement of the system effectiveness to be obtained by extending the system to include the (X, Y) subsystem. Such an upper bound can be expressed as a function of various design configurations R for the (Z, W) subsystem so that a truly documented decision can be made on whether to increase the detail of the world sensed by the system.

6. AN EXAMPLE OF EVALUATION OF ALTERNATIVE INFORMATION SYSTEMS

An analysis of several alternate designs of a hypothetical air defense system is given in this section for three main reasons: 1) to gain an appreciation for potential usefulness of this approach on the future selection and design of information systems, 2) to clarify the theoretical formulations presented in the previous sections, and 3) to illustrate the nature of the calculations which are necessary to carry out the analysis.

Consider the following hypothetical air defense system description as it might be summarized in the system manual.

"Enemy targets are detected by speed and altitude sensors which are located along the defense perimeter. Target altitude and speed information is transmitted from these sensors to the headquarters search station where the information is displayed. On the basis of this information, the target characteristics are determined and the threat is evaluated by the air threat coordinator. The availability of defense forces is then considered, and appropriate weapons are assigned and committed by the air defense commander."

Further technical details of interest are given in later sections of the system manual:

"The information subsystem may be diagrammed as shown in Figure A.

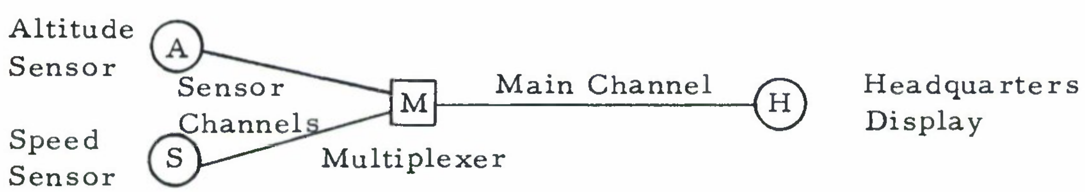


Figure A. The Information System

Altitude and speed information is processed by an analog to digital converter and transmitted to a multiplexer from which it is sent via a common channel to the headquarters display. The altitude and speed sensors operate independently of each other so that the errors of one sensor are statistically independent of the other, i.e., there is no cross talk between them."

Given this brief description of the air defense system, let us determine the value of the information subsystem, i. e., the system diagrammed in Figure A, to the over-all air defense system.

To make the illustrative analysis feasible without the use of a computer, the number of different states of the world and the number of allowable acts open to the air defense commander have been reduced to a number which allows the computations to be made by hand and yet preserves most of the reality of this hypothetical situation.

States of the World and Messages

Enemy missiles, bombers, and light planes are the kind of targets confronting this system. Defensive fighters and missiles will be available to the air defense commander. Let it be assumed that it is sufficient for purposes of classifying enemy targets that the altitude sensor be capable of discriminating three levels: low altitude, medium altitude, and high altitude. It is also assumed that it is sufficient for the speed sensor to distinguish two speeds: low speed and high speed. The state of the world may then be represented by a Boolean vector (a vector whose entries are either one or zero) with three elements, or bits, which are assigned as follows:

$$\{\underbrace{X_1, X_2}_{\substack{\text{altitude} \\ \text{bits}}}, \underbrace{X_3}_{\substack{\text{speed} \\ \text{bit}}}\} = \{X\}$$

These state vectors will be identified by their binary number and are interpreted as follows:

State Vector Number	State Vector Elements		INTERPRETATION	
	$\{X_1, X_2, X_3\}$	Altitude	Speed	Threat
0	0, 0, 0	No Reading	Slow Speed	No Threat
1	0, 0, 1	No Reading	High	No Threat
2	0, 1, 0	Low	Slow	Bomber
3	0, 1, 1	Low	Fast	Light Plane
4	1, 0, 0	Medium	Slow	Bomber
5	1, 0, 1	Medium	Fast	Light Plane
6	1, 1, 0	High	Slow	No Threat
7	1, 1, 1	High	Fast	Missile

The messages about the state of the world generated by the sensors and transmitted to headquarters are also represented by Boolean vectors:

$$\{Y_1, Y_2, Y_3\} = \{Y^j\}$$

in which the i-th element of the j-th message vector, Y_i^j, is the message bit for the corresponding variable of the state vector.

The Action Set, Resources, and Utilities

It will be assumed that the air defense commander will always have at least one fighter and one missile at his disposal, and for every enemy contact he is allowed to take one of the following three actions, denoted by A_i:

A_0 - Do nothing

A_1 - Engage the enemy with a fighter

A_2 - Engage the enemy with a missile

The utility associated with each act and state of the world pair, $u(A_i, X^j)$, must be defined. The utilities can be calculated after the weapon system parameters and the cost and threat values of all weapons are specified.

Let us assume the following values for the system and target parameters:

Cost of one fighter mission	C_{MF}	=	.1
Cost of one defensive missile	C_M	=	1.0
Threat of enemy bomber that penetrates the defense	T_B	=	-20.0
Threat of enemy missile that penetrates the defense	T_M	=	-20.0
Value of destroying enemy bomber	V_B	=	5.0
Value of destroying enemy missile	V_M	=	3.0
Value of destroying enemy light plane	V_F	=	2.0

Probability of kill for fighter against bomber engagement (any altitude)	P_{FB}	=	.75
Probability of kill of fighter against light plane engagement (any altitude)	P_{FF}	=	.5
Probability of kill for missile-bomber engagement for medium altitude bombers	$\overline{P}_{MB}$	=	.7
Probability of kill of missile-light plane engagement for medium altitude light planes	$\overline{P}_{MF}$	=	.7
Probability of kill for missile-missile engagement	P_{MM}	=	.95

The general terms of the utility matrix, $u(A_i, X^j)$, were derived using the above notation and are given as follows:

$$
U = \left[\begin{array}{cccc|cccc}
X^0 & X^1 & X^2 & X^3 & X^4 & X^5 & X^6 & X^7 \\
0 & 0 & T_B & 0 & T_B & 0 & 0 & T_M \\
-C_{MF} & -C_{MF} & \begin{matrix} P_{FB}V_B - C_{MF} \\ (1-P_{FB})\,T_B \end{matrix} & \begin{matrix} P_{FF}V_F \\ -C_{MF} \end{matrix} & \begin{matrix} P_{FB}V_B - C_{MF} \\ (1-P_{FB})\,T_B \end{matrix} & \begin{matrix} P_{FF}V_F \\ -C_{MF} \end{matrix} & -C_{MF} & T_M - C_{MF} \\
-C_M & -C_M & T_B - C_M & -C_M & \begin{matrix} \overline{P}_{MB}V_B - C_M \\ (1-\overline{P}_{MB})\,T_B \end{matrix} & \begin{matrix} \overline{P}_{MF}V_F \\ -C_M \end{matrix} & -C_M & \begin{matrix} V_M - C_M \\ (1-P_{MM})\,T_M \end{matrix}
\end{array} \right]
$$

For example, let's derive $u_{1,4}$, which is the utility of A_1, engaging the enemy with a fighter, where the fourth state of the world is present, i.e., $X^4 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$

which is defined to be a bomber at medium altitude. The $u_{1,4}$ is the sum of the following terms:

The expected value of destroying the bomber with a fighter,

$$P_{FB} V_B$$

The expected threat value of the bomber if the bomber survives the fighter mission,

$$(1-P_{FB})\ T_B$$

The cost of the fighter mission

$$-C_{MF}$$

Therefore, $u_{1,4} = P_{FB}V_B + (1-P_{FB})T_B - C_{MF}$, which is the entry in the second row and fifth column of u_{ij}.

When the values assumed for these terms are substituted, we then have the utility matrix used throughout this analysis:

$$U = \{u_{ij}\} = \begin{bmatrix} 0 & 0 & -20.00 & 0 & -20.00 & 0 & 0 & -20.0 \\ -.1 & -.1 & -1.35 & .9 & -1.35 & .9 & -.1 & -20.1 \\ -1.0 & -1.0 & -21.00 & -1.0 & -3.5 & .4 & -1.0 & 1.0 \end{bmatrix}$$

Every one of the eight possible states of nature, $X^0, X^1, \ldots X^7$, will be assumed to be equally likely. Therefore,

$$\pi = \begin{vmatrix} \frac{1}{8} \\ \frac{1}{8} \\ \cdot \\ \cdot \\ \cdot \\ \frac{1}{8} \end{vmatrix}$$

and therefore,

$$\overline{U} = \frac{1}{8} \begin{bmatrix} 0 & 0 & -20.00 & 0 & -20.0 & 0 & 0 & -20.0 \\ -.1 & -.1 & -1.35 & .9 & -1.35 & .9 & -.- & -20.1 \\ -1.0 & -1.0 & -21.00 & -1 & -3.5 & .4 & -1.0 & 1.0 \end{bmatrix}$$

$\hat{U}$ is given by,

$$\hat{U} = \left[U \pi \right] = \begin{bmatrix} -7.50 \\ -2.68 \\ -3.26 \end{bmatrix} \qquad \text{and}$$

$$\left| \hat{U} \right|^* = -2.68.$$

The Determination of the System Model

The system Q, where $Q = \{p(Y^j | X^i\}$, is the only variable entity which remains to be determined before the system effectiveness, E(Q), may be calculated by

$$E(Q) = \left| \overline{U}\, Q \right|^* \xi - \left| \hat{U} \right|^*$$

The matrix Q may be thought of as a probabilistic model of the information system. It provides a measure of the fidelity with which signals or messages are transmitted; it reflects the degree to which a signal is degraded by noise. Individual components of an information system are influenced by noise in different ways as a result of inherent differences in their functions, design, etc. One must therefore take the Q of each individual subsystem of the information network and properly combine it with the others to get an over-all system Q. The following illustrates how the system model was obtained for one of the system configurations used in this evaluation. The system models for the other configurations are obtained in a similar manner. The system configuration considered is given in Figure B. In this case, the system Q was built up by combining the altitude sensor subsystem, Q_a, the speed sensor subsystem, Q_s, and the main communication link to headquarters, Q_c. Obviously, one could carry the analysis further by taking into consideration the Q of the sensor, of the analog to digital converter, of

the link to the multiplexer, etc. The methods for doing this would remain the same, but, of course, the computational load would increase. The detail to which an analysis is carried would depend on its purpose.

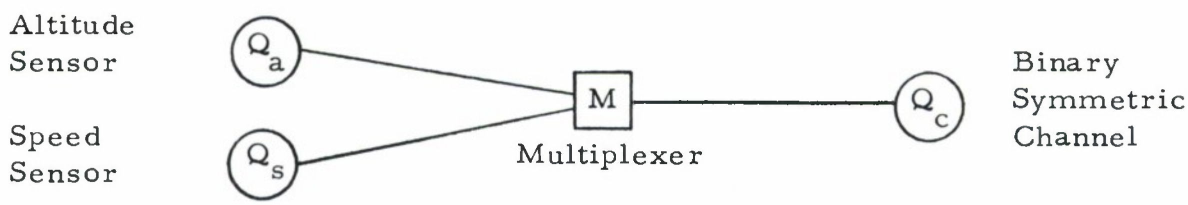


Figure B. Subsystem Models and Their Relations

Altitude and Speed Subsystem

The altitude classification and altitude to signal conversion may be shown as follows, where ρ and λ are specific altitudes with $\rho < \lambda$.

Target Altitude	State of the World	State of the World Vector Representation	
		X_1	X_2
No Target	No Target	0	0
$0 - \rho$	Low Altitude	0	1
$\rho - \lambda$	Medium Altitude	1	0
Above λ	High Altitude	1	1

Let us assume that the only kind of possible error in this subsystem is of the type where the analog to digital conversion may err by only one neighboring unit. Let ρ_a be the probability of this error. Then the possible signals and their associated probability of occurrence resulting from targets at the various altitudes will be as shown in Table 1.

Table 1

Target Altitude	State of the World Vector Representation X_1	X_2	Possible Message Y_1	Y_2	Probability of Message
No Target	0	0	0	0	$1 - \rho_a$
			0	1	ρ_a
			0	0	$\rho_a/2$
Low Altitude	0	1	0	1	$1 - \rho_a$
			1	0	$\rho_a/2$
			0	1	$\rho_a/2$
Medium Altitude	1	0	1	0	$1 - \rho_a$
			1	1	$\rho_a/2$
High Altitude	1	1	1	0	ρ_a
			1	1	$1 - \rho_a$

Using the values from Table 1, it is seen that the altitude subsystem Q matrix is given by:

$$Q_{a_{ij}} = \{p(Y^j|X^i)\} = \begin{bmatrix} 1-\rho_a & \rho_a & 0 & 0 \\ \rho_a/2 & 1-\rho_a & \rho_a/2 & 0 \\ 0 & \rho_a/2 & 1-\rho_a & \rho_a/2 \\ 0 & 0 & \rho_a & 1-\rho_a \end{bmatrix}$$

The above matrix is typical of the noise process associated with the measurement and quantization of a continuous variable.

Let the speed subsystem error rate be denoted by ρ_s. Then the resulting Q matrix is obviously given by

$$Q_{s_{ij}} = \{p_s(Y^j|X^i)\} = \begin{bmatrix} 1-\rho_s & \rho_s \\ \rho_s & 1-\rho_s \end{bmatrix}$$

The Channel Subsystem

The three bits that make up the altitude and speed information are transmitted serially on a binary channel which is assumed to have an error probability r. The error probability is usually small, so that powers higher than the first have been neglected in calculating elements of the Q_c matrix. The Q_c matrix is therefore given by:

$$Q_{c_{ij}} = \{p(Y^j|X^i)\} = \begin{matrix} 1-3r & r & r & 0 & r & 0 & 0 & 0 \\ r & 1-3r & 0 & r & 0 & r & 0 & 0 \\ r & 0 & 1-3r & r & 0 & 0 & r & 0 \\ 0 & r & r & 1-3r & 0 & 0 & 0 & r \\ r & 0 & 0 & 0 & 1-3r & r & r & 0 \\ 0 & r & 0 & 0 & r & 1-3r & 0 & r \\ 0 & 0 & r & 0 & r & 0 & 1-3r & r \\ 0 & 0 & 0 & r & 0 & r & r & 1-3r \end{matrix}$$

Typical elements of Q_c are calculated as follows:

$$p(Y^2|X^2) = p(010\,|\,010) = (1-r)(1-r)(1-r) = (1-r)^3$$

$$= 1-3r \text{ when neglecting high order terms}$$

$$p(Y^4|X^2) = p(100\,|010) = r\,r(1-r) = r^2 - r^3 = 0$$

System Q

The procedure for finding the system Q applies the same kind of approach used in finding an over-all equivalent impedance of an electrical network. In an electric circuit one does this by combining individual impedances in various series and parallel combinations according to the rules of physics and vector calculus. In our case the rules for finding system Q follow from the laws of probability and matrix algebra. We shall achieve the combination in the following order:

1) Q_s and Q_a are combined to form an equivalent Q_{sensor}, i.e.,

$$\left.\begin{matrix} Q_a \\ \\ Q_s \end{matrix}\right\rangle Q_c \;\text{———}\; \equiv \;\text{———}\; Q_{sensor} \;\text{———}\; Q_c \;\text{———}$$

2) Q_{sensor} is combined with Q_c to form the equivalent system Q, i.e.,

$$Q_{sensor} \;\text{———}\; Q_c \;\text{———}\; \equiv \;\text{———}\; Q_{system} \;\text{———}$$

In our system, Q_{sensor} is an 8 x 8 matrix which shows the effect of noise introduced in the sensor subsystems. The individual elements of Q_{sensor} are joint probabilities, and because of the assumed statistical independence of the operation of the altitude and speed subsystems, they are calculated as follows:

$$Q_{sensor \atop jkl,\, mnp} = \{p_{sensor}(Y^{jkl}|X^{mnp})\} = p_a(Y^{jk}|X^{mn})\, p_s(Y^{l}|X^{p})$$

For example,

$$p_{sensor}(Y^{000}|X^{000}) = p_a(Y^{00}|X^{00})\, p_s(Y^{0}|X^{0}) = (1-\rho_a)(1-\rho_s)$$

or,

$$p_{sensor}(Y^{001}|X^{000}) = p_a(Y^{00}|X^{00})\ p_s(Y^1|X^0) = (1-\rho_a)\,\rho_s$$

or,

$$p_{sensor}(Y^{001}|X^{010}) = p_a(Y^{00}|X^{01})\ p_s(Y^1|X^0) = \frac{\rho_a}{2}\ \rho_s$$

etc.

The information system is now reduced to a pair of Q's in series which may be combined by matrix multiplication. That is,

$$\{p(Y^j|X^i)\} = Q = Q_{sensor}\ Q_c$$

The reason for this operation becomes clear when one examines the terms of any element of Q. Take the term Q(2, 2), for example, which is $p(Y^{010}|X^{010})$. According to the matrix multiplication indicated above, this term is equal to

$$p(Y^{010}|X^{010}) = \sum_{j=0}^{7} p_{sensor}(Z^j|X^{010})\ p_c(Y^{010}|Z^j)$$

where Z^j is the j-th output message of the sensor subsystem which acts as the j-th input message for the communication subsystem. The above is the correct expression for the probability of the message Y^{010} given that the state of the world is X^{010}. This is exactly what this element of the Q_s matrix is supposed to represent.

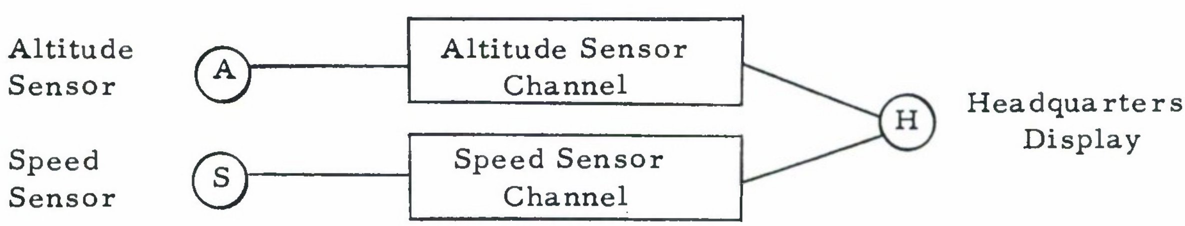
6.1 Alternative Configurations

In the following pages five sets of systems assumptions are evaluated. The assumptions were selected mainly with the purpose of making a number of points which indicate what can be obtained from evaluations of this type. It is important to point out that the essential characteristic of the design process is the judicious choice of a sequence of design assumptions. The method does not entirely determine such choices but guides them by showing the relative merits of alternate choices. It is not easy, of course, to render the flavor of a true design effort by using a few sets of assumptions, but the points made in the discussion should tend to do so in a partial manner.

6.2 The Determination of Channel Requirements for Various Sensor Inputs

Let us assume that one has n variables being sensed by the sensors $S_1, S_2, \ldots S_n$. An important question is what should be the amount of noise to be tolerated for each signal to obtain maximum effectiveness for a fixed cost (or maybe maximum effectiveness per dollar). If, in particular, one assumes that the noise is due mainly to the channel on which the sensor signals are to be sent, such a question can be answered by assuming a system configuration of the following type:*

System Assumption # 1



- Altitude Sensor - perfect, no noise ($Q_A = I$)
- Speed Sensor - perfect, no noise ($Q_{SP} = I$)
- Altitude sensor channel - binary, symmetric, noisy with error rate r_a.

---

* A similar analysis is possible if the sensors are also assumed to be noisy.

. Speed sensor channel - binary, symmetric, noisy with error rate r_s.

. Double and triple errors are neglected in computing $Q_C = Q_A * Q_{SP}$.

The Q of the system will be function of the parameters r_a and r_s. Thus, the system effectiveness is plotted on the r_a r_s plane in Figures 1 and 2. In Figure 1 the regions where distinct decision rules prevail are indicated; in Figure 2 the lines of equal effectiveness or isoquants are given. Also in Figure 2 an iso-cost line for the cost assumption $r_a\ r_s = K = .01$ is indicated. It should be pointed out that the (.5, .5) point corresponds to the null system, i.e., the system which provides headquarters with no information whatever about the environment. The point (0, 0) corresponds to a system which makes no errors providing headquarters with perfect knowledge of the environment. The system designer could optimize the design by seeing where along this iso-cost curve system effectiveness is maximized. In System 1 effectiveness is maximized at $r_s = .5$ and $r_a = .02$. This is an extreme case and suggests that at this particular level of funding the speed sensor is not necessary. The extreme location of this solution is a result of the approximations made in finding $Q_C = Q_A * Q_{SP}$. Actually, if the higher order terms had not been neglected, the effectiveness would have decreased more rapidly than is indicated in Figure 2, and the optimum point would probably have been located near $r_s = .35$, $r_a = .03$.

Demonstrated here is the fact that the most effective fixed cost system is not one with equal error rates for the two signals, but one in which the error rates are very uneven. Without an analysis utilizing an underlying structure of the entire system, one would have no rationale to make such an assertion. Moreover, one would probably have adopted an equal accuracy design, possibly based on such an irrelevant criterion as the minimum sum of errors function for a given allocation of funds.

In general, one attempts to use a single communication channel by feeding the output of various sensors into a time or frequency multiplexer.

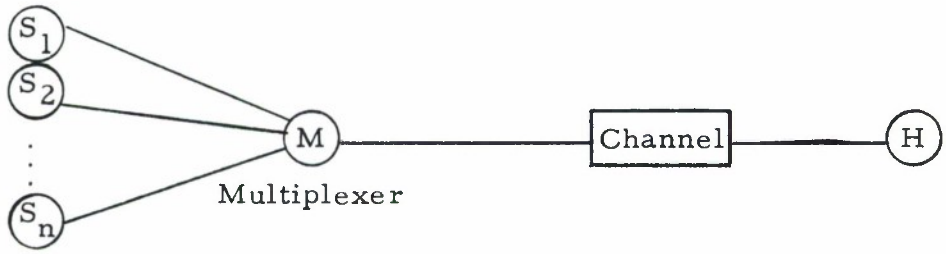


Figure C

Thus, one might object that the system assumption given above is not a useful one, since it presupposes separate channels for the altitude and speed signals. The point to be made is that such an assumption should be considered as a prerequisite for assessing if multiplexing can be performed without any encoding of the various signals. What we have in mind is that if the optimal choices for the error probabilities of the various individual channels turn out to be nearly the same, one can multiplex the signals and feed them to a common channel without further ado. The capacity of such channels would be determined 2 by the bit rate and the said (common) error probability. If, on the other hand, the various recommendations of the error rates are highly dissimilar, one has to encode the signals for which high accuracy is required so that their encoded form can be transmitted at the noise levels which are permissible for the lowest accuracy signal. The common channel can then be designed for the lowest accuracy, thereby avoiding the waste of channel capacity. Thus, in the case of this example, the system to be realized would take the form illustrated in Figure D.

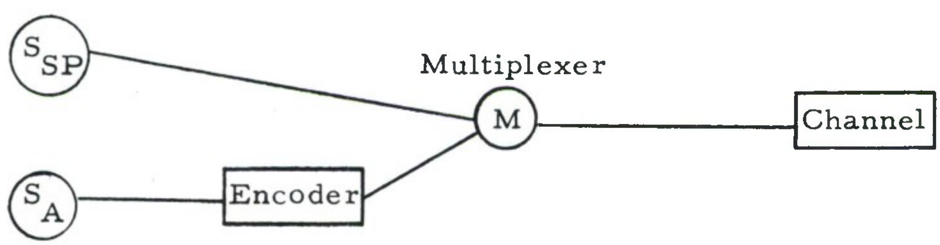


Figure D

<u>System Assumption # 2</u>

This is identical to the above system assumption except that the altitude signal is assumed to be coded with a single error correcting code; in other words, the Q_A part of the system model

$$Q_C = Q_{SP} * Q_A$$

is given by:

$$
\begin{array}{cc}
 & Y^j \\
 & \begin{array}{cccc} 00 & 01 & 10 & 11 \end{array} \\
X^i \begin{array}{c} 00 \\ 01 \\ 10 \\ 11 \end{array} & \left|\begin{array}{cccc}
1-R_a & 0 & 0 & R_a \\
0 & 1-R_a & R_a & 0 \\
0 & R_a & 1-R_a & 0 \\
R_a & 0 & 0 & 1-R_a
\end{array}\right|
\end{array}
$$

instead of:

$$
\begin{array}{cc}
 & Y^j \\
 & \begin{array}{cccc} 00 & 01 & 10 & 11 \end{array} \\
X^i \begin{array}{c} 00 \\ 01 \\ 10 \\ 11 \end{array} & \left|\begin{array}{cccc}
(1-r_a)^2 & r_a(1-r_a) & r_a(1-r_a) & r_a^{\,2} \\
r_a(1-r_a) & (1-r_a)^2 & r_a^{\,2} & r_a(1-r_a) \\
r_a(1-r_a) & r_a^{\,2} & (1-r_a)^2 & r_a(1-r_a) \\
r_a^{\,2} & r_a(1-r_a) & r_a(1-r_a) & (1-r_a)^2
\end{array}\right|
\end{array}
$$

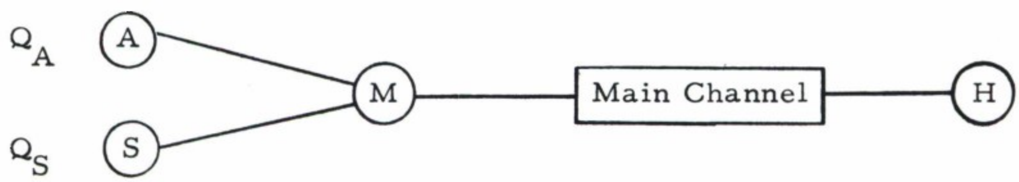
R_a is the probability of a double error, and r_a is the probability of a single error.

A comparison of the isoquants for Systems 1 and 2, Figures 2 and 4, brings out another interesting effect, namely, the unexpected manner in which coding can change the dependence of system effectiveness on the error rates of the channels. The comparison of these two systems shows that the criticality of the speed information is decreased if the information in the altitude information is single error corrected. Again, this sort of effect can be determined only if the interaction of the various design parameters is considered through a utility structure.

6.3 Sensor versus Channel Accuracies

The next set of three systems assumptions is introduced for two reasons: The first is to introduce systems with both parallel and serial connections of subsystems. The second is to show how the over-all system model permits one to balance out the design of all the components of the system. That is, one can find an optimal set of accuracy specifications for the sensors, channels, buffers, etc., of the system so that such specification is internally consistent. The specifications of these systems are:

System 3. This is the general configuration for which the detailed computations were presented in the previous few pages. A diagram of this system is given here again.



The system assumptions are:

Altitude sensor	-	noisy with error rate to neighboring unit ρ_a
Speed sensor	-	noisy with error rate ρ_s
Sensor channels	-	perfect, no noise, $r = 0$ ($Q_{SC} = I$)
Main channel	-	binary, symmetric, noisy with error rate $r = .1$. High order terms are neglected in calculating Q_C.

System 4. The same system as System 3 except for the main channel. The main channel is noisy with an error rate $r = .1$. Single errors are eliminated by the adoption of a single error correcting code.

System 5. The same system as System 3 except for the main channel. The main channel is assumed to be perfect, $r = 0$ ($Q_C = I$).

The utility matrix, the state distribution, and the allowable acts remain the same for Systems 1 through 5. The detailed computations for Systems 1 through 4 are not presented since the mathematics is essentially the same as for System 5.

Comparing Systems 3, 4, and 5, Figures 6, 8, and 10, it is immediately apparent that System 5 is uniformly better than System 4, which in turn is uniformly better than System 3. This is as it should be, since going from System 5 to System 3 the communication channel becomes increasingly more noisy while the sensing subsystem is the same for all three. Comparing Figures 6, 8, and 10 also shows that most of the effectiveness loss due to degrading the channels from a perfect system (System 5) to a system with a 10% error rate (System 3) is recouped by the introduction of a single error correcting channel (System 4). In other words, in this case the adoption of a single error correcting code for the main communication link to headquarters results in essentially neutralizing the effects of a 10% error rate in this link. It should be pointed out that a 10% error rate is an extremely high one for communication channels and would actually result from man-made interference and jamming. Here we see in a simple way how this type of evaluation can furnish quantitative arguments for the adoption of a coding scheme for a channel which one can expect to be subjected to man-made interference.

The sensors noise for Systems 3, 4, and 5 have been characterized by the matrices:

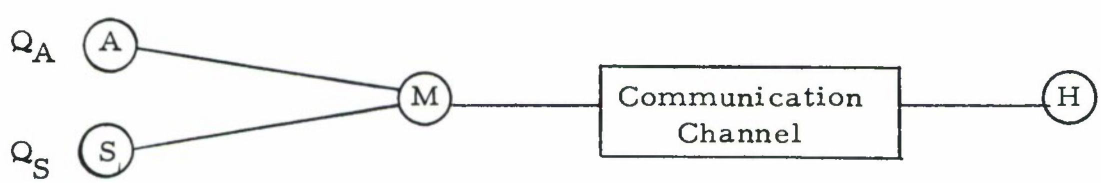
$$\begin{array}{cc|cccc|} & & & & Y_a^j & \\ & & 00 & 01 & 10 & 11 \\ \hline & 00 & (1-\rho_a) & \rho_a & 0 & 0 \\ X_a^i & 01 & 1/2 \;\; _a & (1-\rho_a) & 1/2\,\rho_a & 0 \\ & 10 & 0 & 1/2\,\rho_a & (1-\rho_a) & 1/2\,\rho_a \\ & 11 & 0 & 0 & \rho_a & (1-\rho_a) \end{array}$$

for the altitude sensor, and by

$$
\begin{array}{ccc}
 & & Y_s^j \\
 & 0 & 1 \\
X_s^i \begin{array}{c} 0 \\ 1 \end{array} & \multicolumn{2}{c}{\left|\begin{array}{cc} 1-\rho_s & \rho_s \\ \rho_s & 1-\rho_s \end{array}\right|}
\end{array}
$$

for the speed sensor. Sensing errors can occur only as jumpts to an adjacent value of the variable being measured. This type of error process is characteristic of the measurement of an analog quantity. It would also be characteristic of a communication channel utilizing non-digitalized signals as, for example, a radio link utilizing for amplitude modulation envelope the value of an analog quantity. On the other hand, if a digitalized representation is used (a representation in which the position in the code block is significant), the channel noise process is quite different, since errors can occur between non-adjacent values. Because of these characteristics of digitalized transmission, errors in a digital channel can be much more harmful than those in an analog channel or than those in the sensors which are usually of analog type. The simple examples we have computed demonstrate this in a quantitative fashion, at least with regard to the relative importance between the sensor noise and the digital channel noise.

We can schematically represent Systems 3, 4, and 5 as follows:



The point of coordinates (0, 0) in Figure 10 gives the effectiveness of the system which has noise-free sensors and channels. This effectiveness is E(Q) = 2.65 and is the least upper bound for <u>all</u> systems with the above structure. Let us examine what happens to the effectiveness of the system as we degrade its three components from the 0% to the 10% error rate level.

a) Degrade the speed sensor (0% $\longrightarrow$ 10%) but leave the altitude sensor and the channel at the no-error level. E(Q) is read at the point of coordinates $\rho_s = .1$ $\rho_a = 0$ of Figure 10; it is seen to be E(Q) = 2.19. The performance of the system has lost about 6% of perfect performance.

b) Degrade the altitude sensor by 10% leaving the speed sensor and channel perfect. This corresponds to the point $\rho_a = .1$ $\rho_s = 0$ of Figure 10. Then E(Q) = 2.54 with a loss of about 4% of perfect performance.

c) Degrade both altitude and speed sensor by 10% leaving the channel perfect, $\rho_a = .1$ $\rho_s = .1$ in Figure 10. E(Q) = 2.36 or about 11% of perfect performance is lost.

d) Degrade the channel by 10% but leave the two sensors perfect. This corresponds to the point $\rho_a = 0$ $\rho_s = 0$ of Figure 6 (which corresponds to the system with the channel with an error rate of 10%). E(Q) = 1.97. <u>This is a loss of better than 25% of the perfect performance.</u>

Thus, we see that a 10% degradation of the communication channel is much more harmful than a similar degradation for <u>all the sensors.</u>

The above comparisons indicate then that an optimal design may require a greater accuracy for the channel than for the sensors and not an equal one. Furthermore, one has obtained a quantitative measure concerning such relative accuracies. These are not simple orderings derived through intuition.

Observing the manner in which the shape of the isoquants change in going from System 3 to 4 to 5, one sees the strong influence which coding of the signals in the main channel has on changing the relative importance of the speed and altitude sensor errors on system effectiveness. It is seen that the criticality of the speed sensor is progressively reduced in advancing to the system with the perfect channel. This indicates that if the channel is jammed, one has to rely more heavily on the speed of the threat information to discriminate missiles from planes.

6.4 Decision Rules

The set of figures 1, 3, 5, 7, and 9 indicate the way the indifference lines partition the system space. An indifference line or contour is a locus of those system designs which admit at least two optimal decision rules. A decision rule is a set of prescriptions of the type $Y^j \longrightarrow A_k$ which can be read: if Y^j is observed, select the action A_k. The indifference lines are then the locus of points which divide the system designs into those which admit as the optimal rule some $A_k = d_1(Y^j)$ and those which admit some $A_k = d_2(Y^j)$. The two rules d_1 and d_2 differ in at least one prescription $Y^j \longrightarrow A_k$. If d_1 and d_2 differ in the prescription $Y^j \longrightarrow A_k$, the indifference line is marked with the said Y^j and also with a marking which shows which A_k is to be selected on each side of the indifference line. As an example, consider the region near the origin of Figure 9. This is the region in which both the sensors and channels are nearly perfect. The decision rule is seen to be:

$$
\begin{array}{llll}
Y^0 \longrightarrow A_0 & & & Y^4 \longrightarrow A_1 \\
Y^1 \longrightarrow A_0 & & & Y^5 \longrightarrow A_1 \\
Y^2 \longrightarrow A_1 & & & Y^6 \longrightarrow A_0 \\
Y^3 \longrightarrow A_1 & & & Y^7 \longrightarrow A_2
\end{array}
$$

Because the systems with design characteristics within that region are systems that have infrequent errors, we can say that if we receive a message indicating that there is a bomber, we can be almost totally sure that in actuality we are facing a bomber. Thus, the decision rule is:

$Y^0 \equiv$ No Threat $\longrightarrow$ Do Nothing $\equiv A_0$
$Y^1 \equiv$ No Threat $\longrightarrow$ Do Nothing $\equiv A_0$
$Y^2 \equiv$ Low Altitude Bomber $\longrightarrow$ Engage with Fighter $\equiv A_1$
$Y^3 \equiv$ Low Altitude Light Plane $\longrightarrow$ Engage with Fighter $\equiv A_1$
$Y^4 \equiv$ High Altitude Bomber $\longrightarrow$ Engage with Fighter $\equiv A_1$
$Y^5 \equiv$ High Altitude Light Plane $\longrightarrow$ Engage with Fighter $\equiv A_1$
$Y^6 \equiv$ No Threat $\longrightarrow$ Do Nothing $\equiv A_0$
$Y^7 \equiv$ Missile $\longrightarrow$ Engage with Missile $\equiv A_2$

which of course is the decision rule that intuition suggests to be best. If, on the other hand, we are receiving information from the system corresponding to the $\rho_a = .1$ $\rho_s = .1$ point of Figure 5, i.e., if because of enemy action our sensors and channels both have an error rate of 10%, due to jamming or actual physical damage, we can trust our system much less, and we escalate our response. In fact, the optimal decision rule is:

$Y^0 \equiv$ "No Threat" $\longrightarrow A_1 \equiv$ Engage with Fighter

$Y^1 \equiv$ "No Threat" $\longrightarrow A_1 \equiv$ Engage with Fighter

$Y^2 \equiv$ "Low Altitude Bomber" $\longrightarrow A_1 \equiv$ Engage with Fighter

$Y^3 \equiv$ "Low Altitude Light Plane" $\longrightarrow A_1 \equiv$ Engage with Fighter

$Y^4 \equiv$ "High Altitude Bomber" $\longrightarrow A_1 \equiv$ Engage with Fighter

$Y^5 \equiv$ "High Altitude Light Plane" $\longrightarrow A_2 \equiv$ Engage with Missile

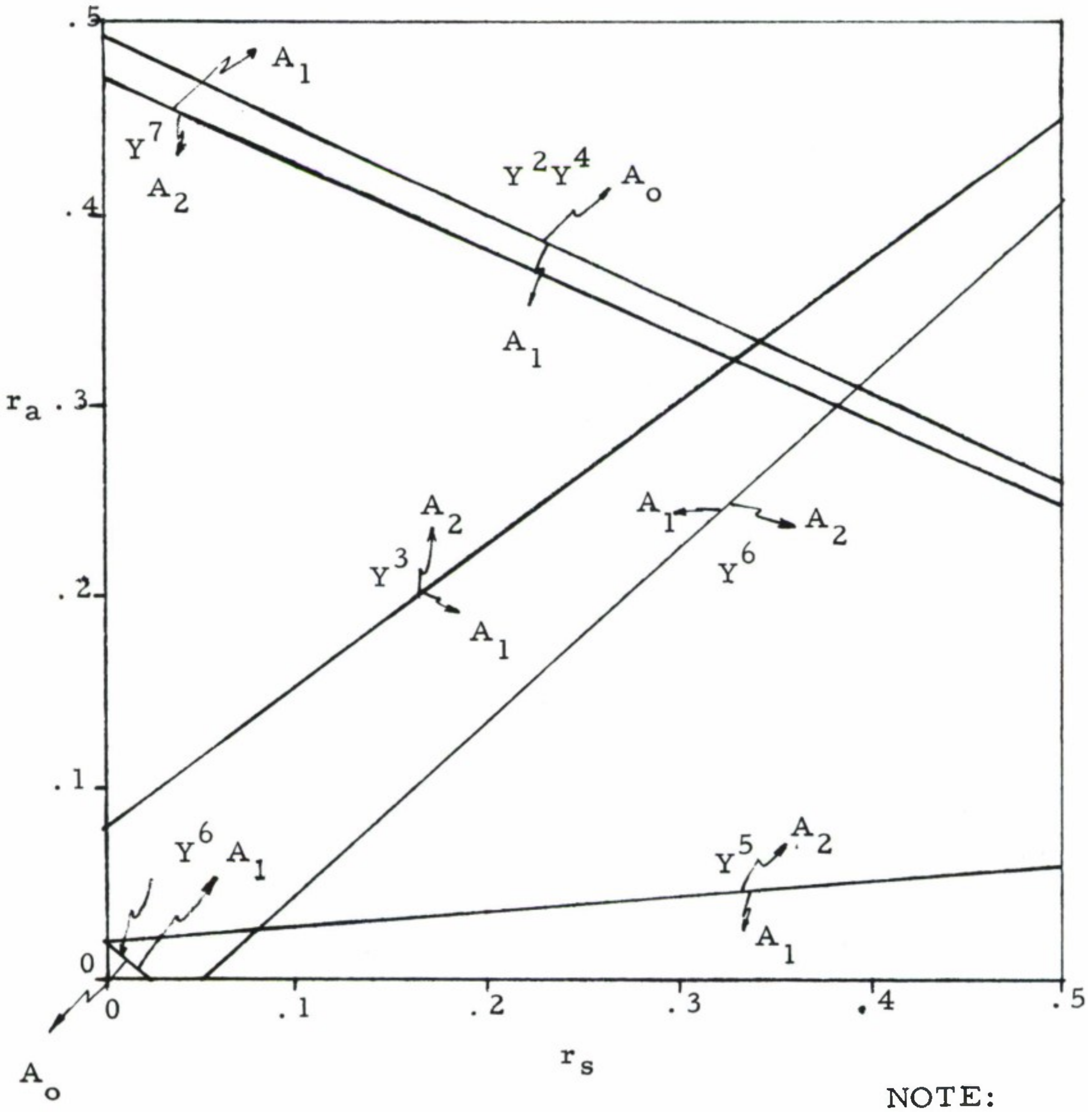
$Y^6 \equiv$ "No Threat" $\longrightarrow A_2 \equiv$ Engage with Missile

$Y^7 \equiv$ "Missile" $\longrightarrow A_2 \equiv$ Engage with Missile

The various interpretations of the signals are subject to doubt due to relatively high noise content in the system. This is why Y^0 is indicated to have a "No Threat" interpretation where the quotation marks emphasize that Y^0 may correspond to other things. For example, $Y^0 = 000$ may have originated from $X^2 = 010$ which is a bomber. This is why a supposedly no threat condition evokes a fighter mission response as insurance against the errors of the information system. The cases of Y^5 and Y^6 are particularly interesting for $Y^5 = 101 \equiv$ light plane. An error in the second altitude digit (due to either the channel or the sensor) may have caused $Y^5 = 101$ to actually originate from $X^7 = 111$, that is, a missile. This is the main reason for responding then with an anti-missile-missile. For $Y^6 = 110$, which is a no threat condition, an error in the speed bit could cause a missile not to be detected. This is why an anti-missile-missile response is advisable. By comparing the two decision rules so obtained, one can clearly see that a 10% degradation of the three components of the system causes a major escalation of the responses specified by the optimal decision rules.

SYSTEM 1

DECISION RULES $d(r_s, r_a)$



Constant Decision Rules

Message	Action
Y^0	A_1
Y^1	A_1

The lines shown indicate the boundaries which divide regions requiring a different rule of response. Each line is labeled with a) the message requiring the differential response, b) the optimal response on both sides of the line.

NOTE:

A_0 : Do Nothing
A_1 : Engage with fighter
A_2 : Engage with missile
Y^j : Defined on page 27

Figure 1. Decision Rules.

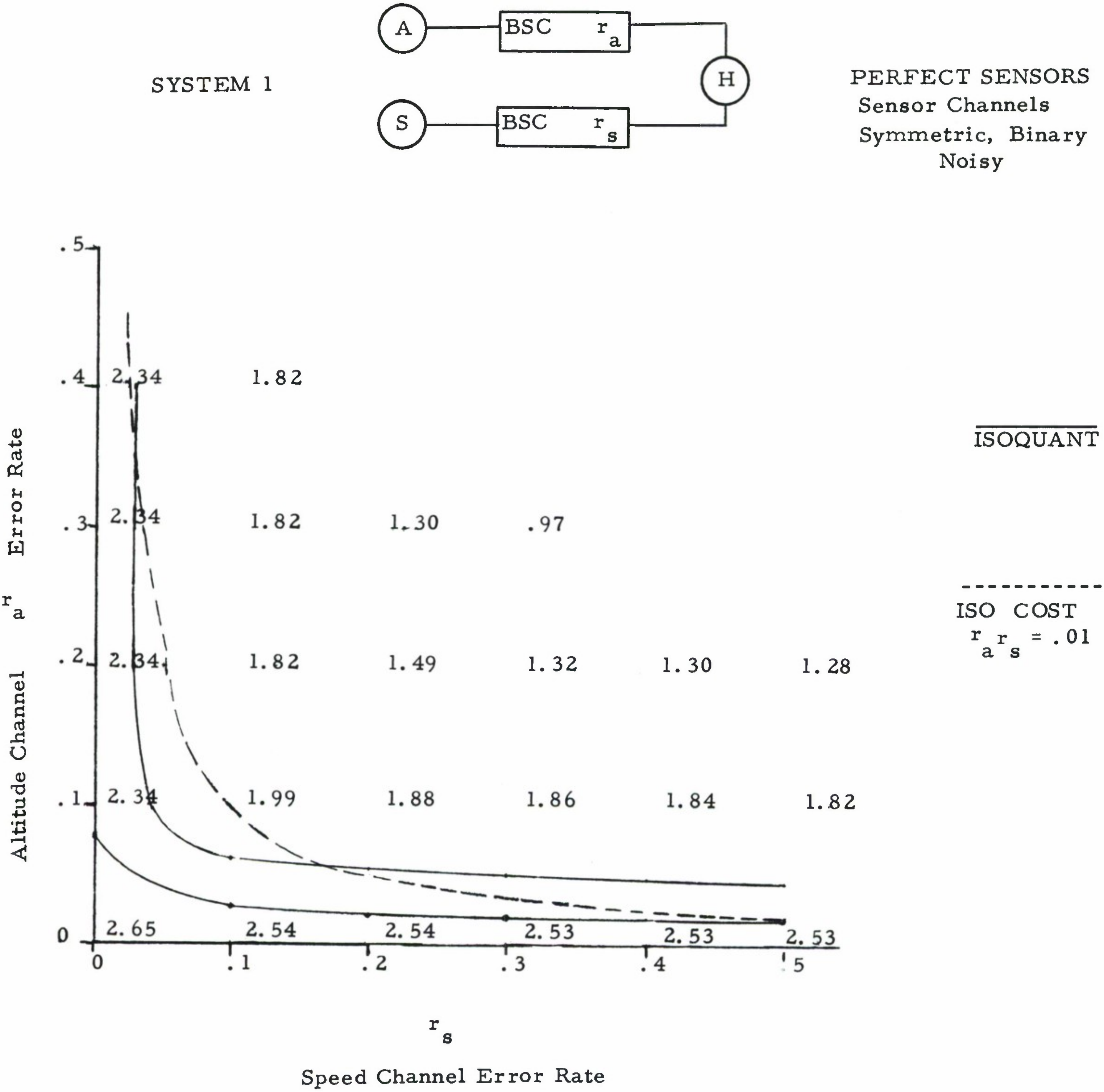
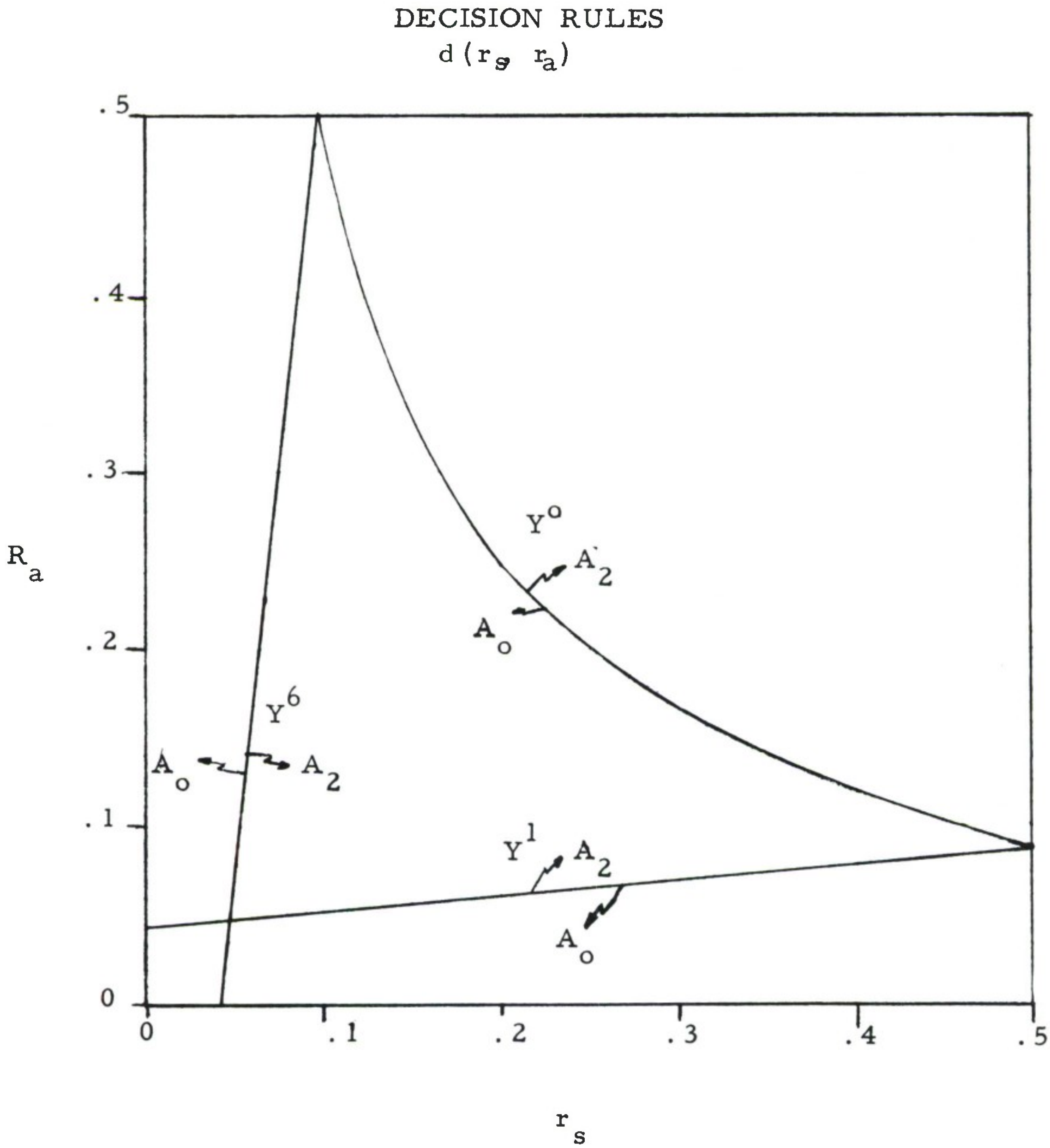


Figure 2. System Effectiveness, E (Q)



Constant Decision Rules

Message	Action
Y^2	A_1
Y^3	A_1
Y^4	A_1
Y^5	A_1
Y^7	A_2

The lines shown indicate the boundaries which divide regions requiring a different rule of response. Each line is labeled with a) the message requiring the differential response, b) the optimal response on both sides of the line.

NOTE:
A_o: Do Nothing
A_1: Engage with fighter
A_2: Engage with missile
Y^j: Defined on Page 27

Figure 3. Decision Rules

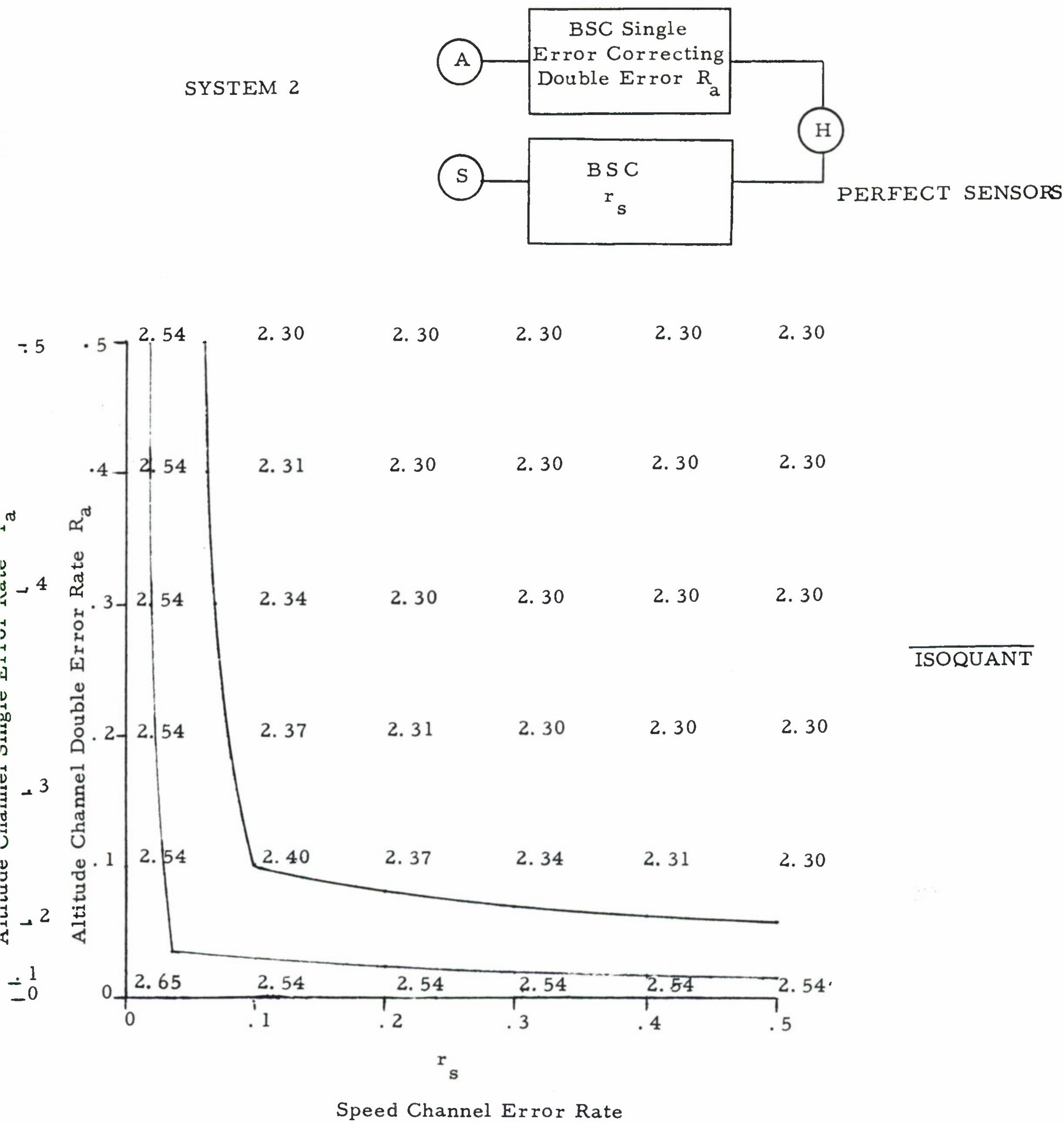
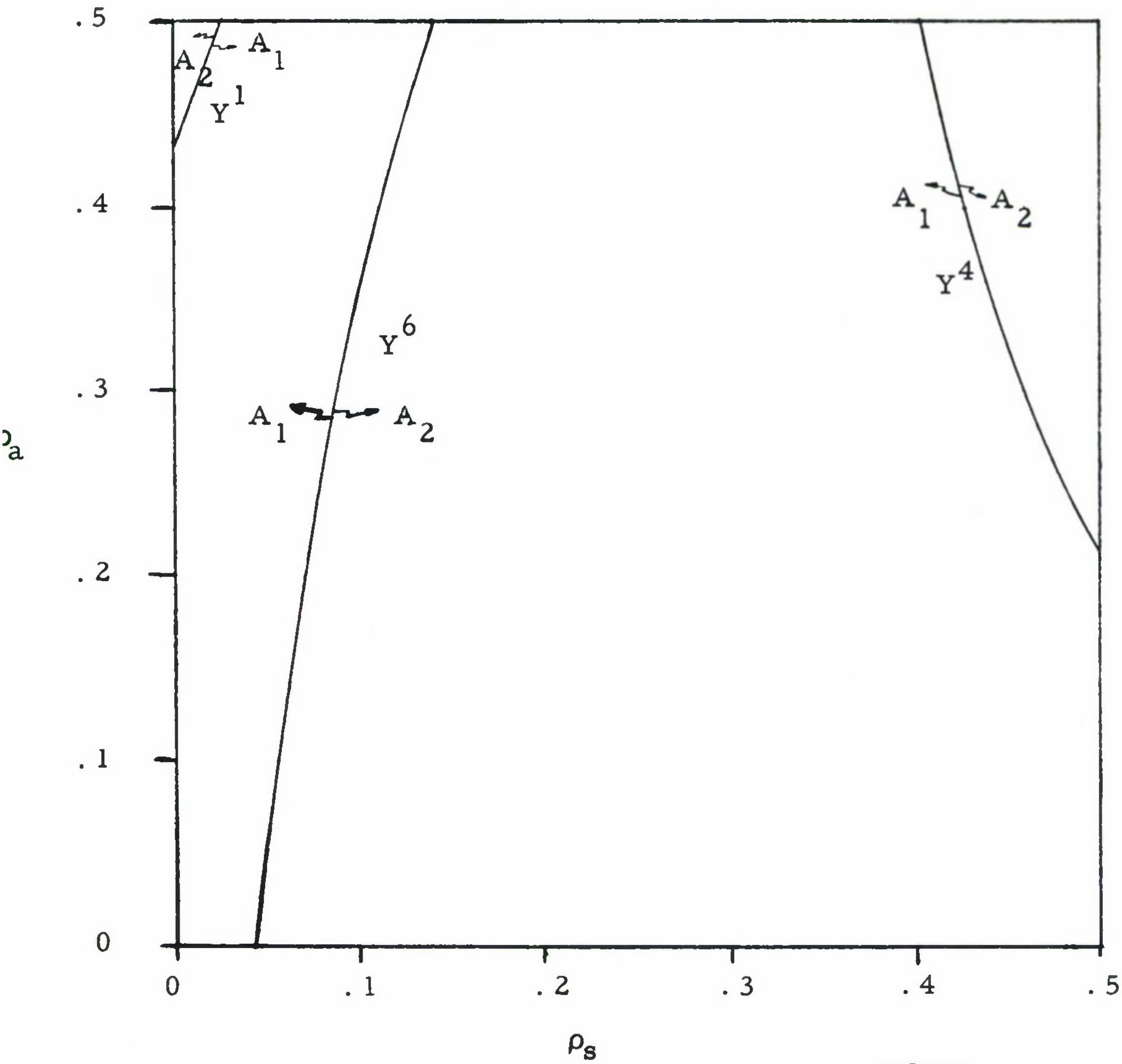


Figure 4. System Effectiveness, E(Q)

DECISION RULES $d(\rho_s, \rho_a)$



Constant Decision Rules

Message	Action
Y^o	A_1
Y^2	A_1
Y^3	A_1
Y^5	A_2
Y^7	A_2

The lines shown indicate the boundaries which divide regions requiring a different rule of response. Each line is labeled with a) the message requiring the differential response, b) the optimal response on both sides of the line.

NOTE:

A_o: Do nothing

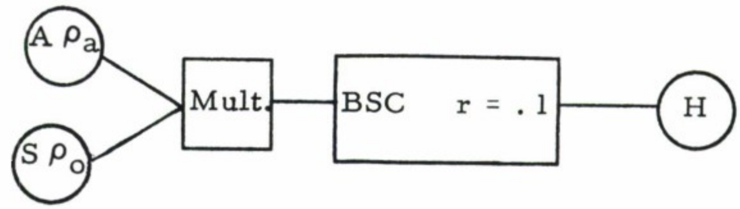
A_1: Engage with fighter

A_2: Engage with missile

Y^j: Defined on page 27

Figure 5. Decision Rules.

SYSTEM 3



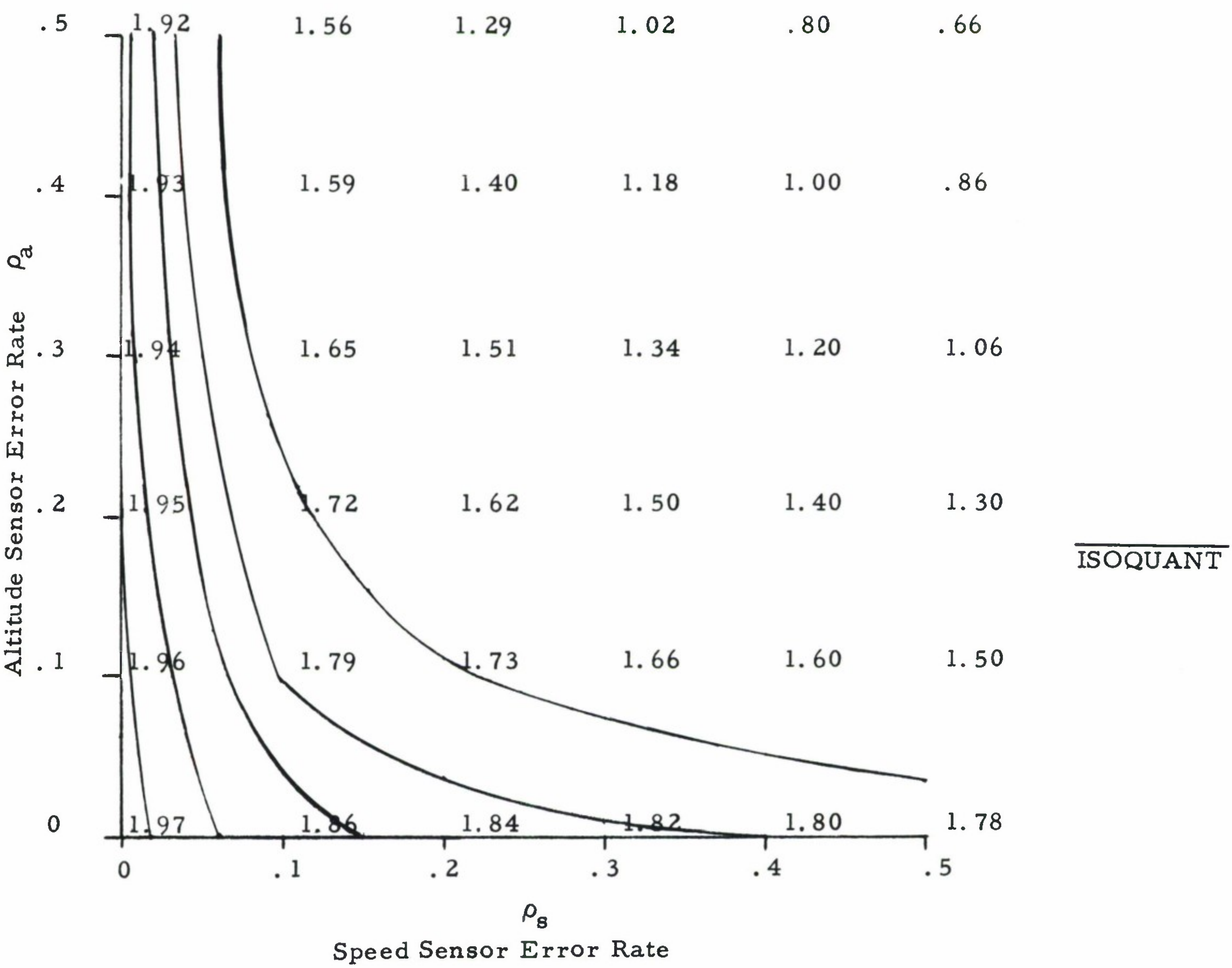
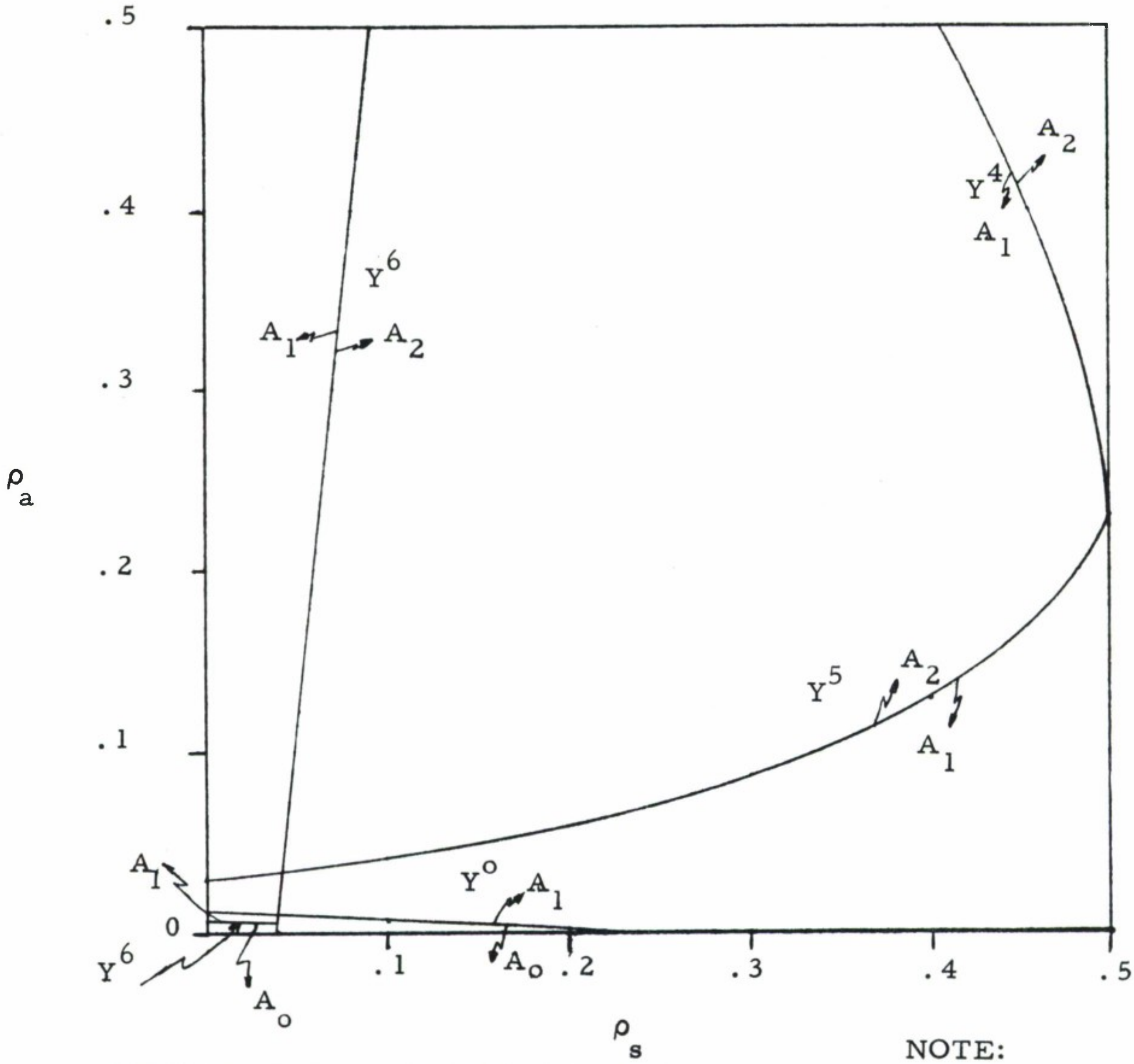


Figure 6. System Effectiveness, E(Q)

DECISION RULES $d(\rho_s, \rho_a)$



Constant Decision Rules

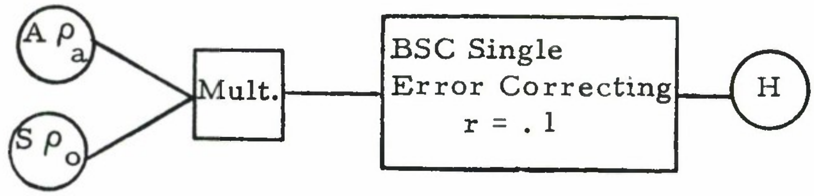
Message	Action
Y^1	A_1
Y^2	A_1
Y^3	A_1
Y^7	A_2

The lines shown indicate the boundaries which divide regions requiring a different rule of response. Each line is labeled with a) the message requiring the differential response, b) the optimal response on both sides of the line.

NOTE:
A_o: Do nothing
A_1: Engage with fighter
A_2: Engage with missile
Yj: Defined on page 27

Figure 7. Decision Rules.

SYSTEM 4



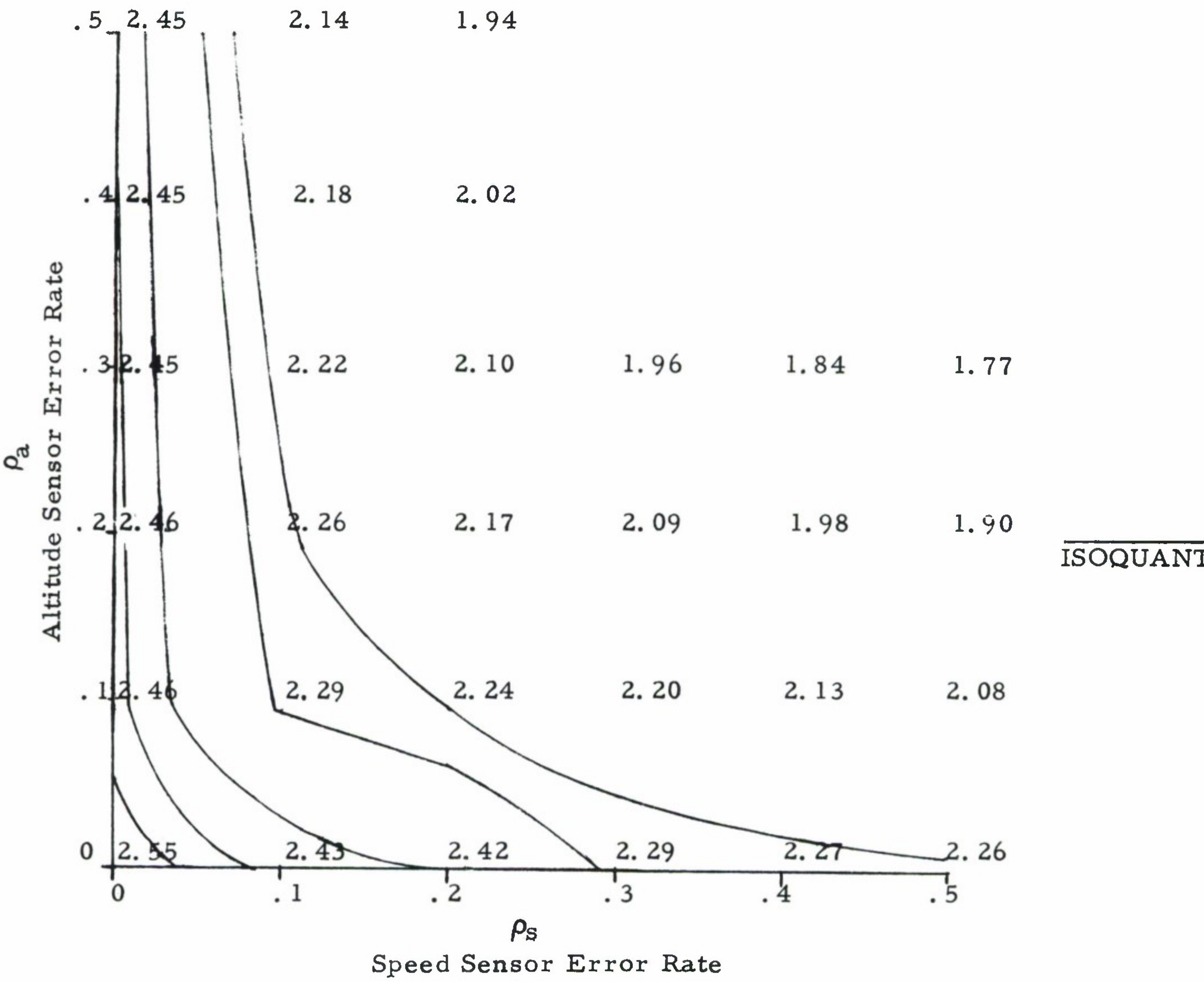
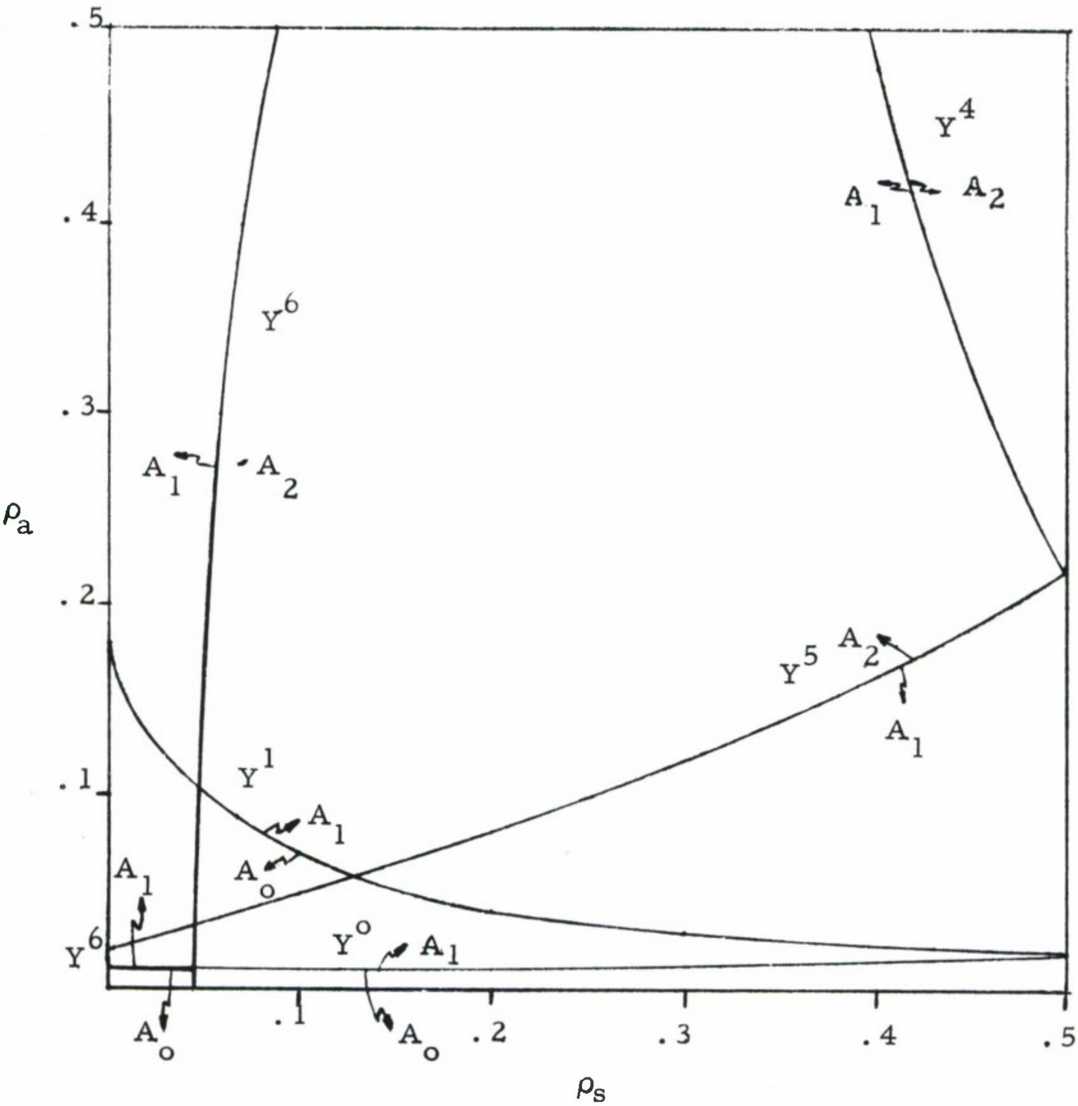


Figure 8. System Effectiveness, E(Q)

DECISION RULE $d(\rho_s, \rho_a)$



Constant Decision Rules

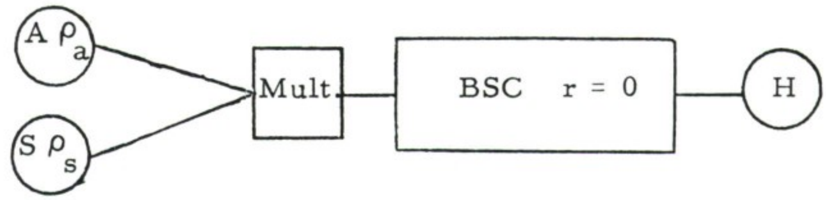
Message	Action
Y^2	A_1
Y^3	A_1
Y^7	A_2

The lines shown indicate the boundaries which divide regions requiring a different rule of response. Each line is labeled with a) the message requiring the differential response, b) the optimal response on both sides of the line.

NOTE:
- A_o: Do nothing
- A_1: Engage with fighter
- A_2: Engage with missile
- Y^j: Defined on page 27

Figure 9. Decision Rules

SYSTEM 5



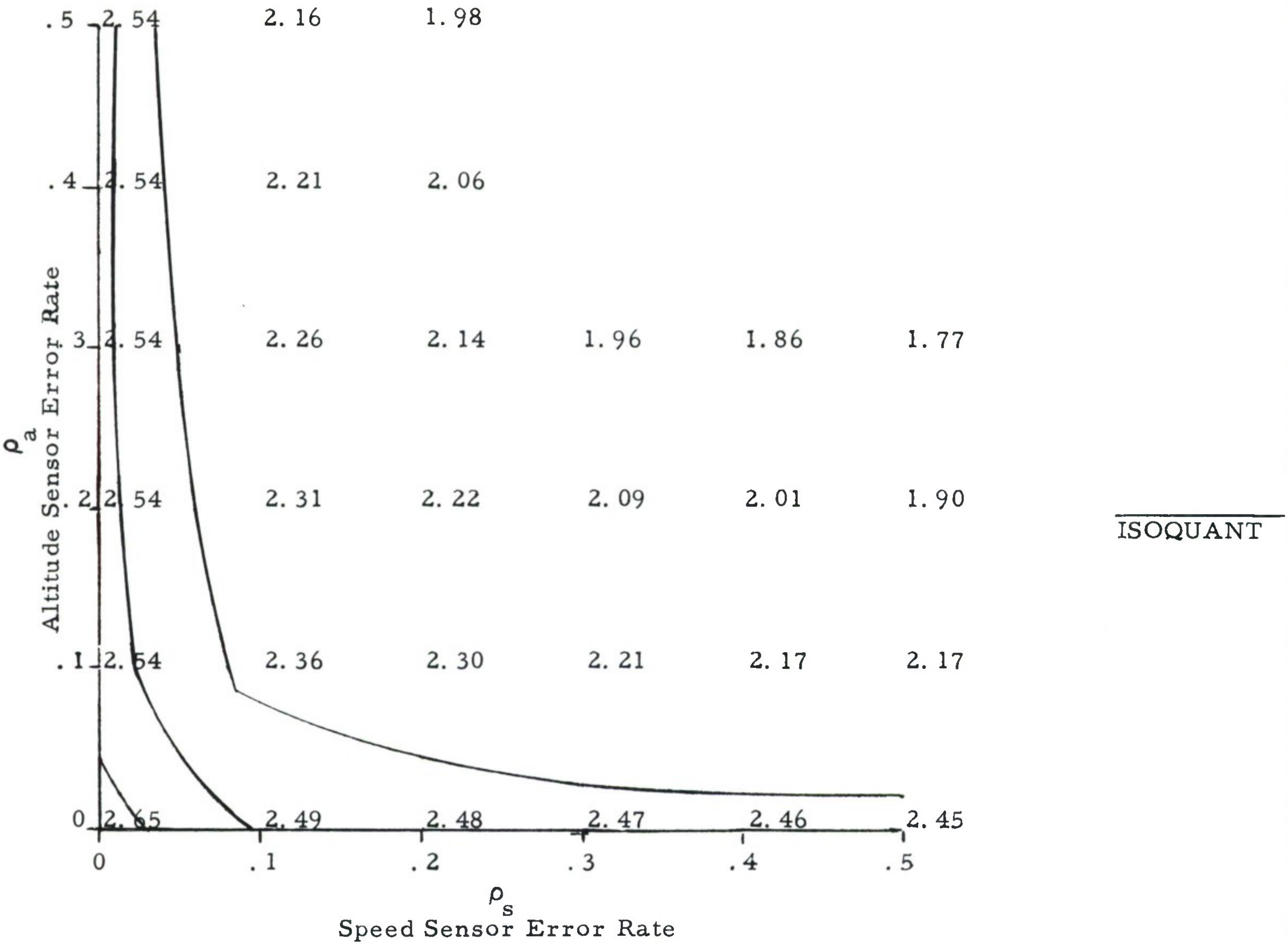


Figure 10. System Effectiveness, E(Q)

REFERENCES

1. Raiffa, H., and Schlaifer, R., Applied Statistical Decision Theory, Harvard Business School, Division of Research, Boston, 1961.

2. Kuhn, H.W., and Tucker, A.W., "Non Linear Programming," Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Edited by Jergy Newman, University of California Press, Berkeley and Los Angeles, 1951.

3. Shannon, C. "A Mathematical Theory of Communication," Bell System Technical Journal, 1948.

4. Kelly, J. "A New Interpretation of Information Rate," Bell System Technical Journal, 1956.

5. Marschak, J. "Remarks on the Economics of Information," Contribution to Scientific Research in Management, 1960.

APPENDIX I

1. AN IDEALIZED BOMB-DAMAGE-ASSESSMENT COMMAND AND CONTROL SYSTEM

1.1 Description of the Organization

Assume that a military organization is engaged in the activity of bombing a set of n enemy targets. After the initial round of bombing, the organization must make decisions on how to allocate the remaining weapons on the basis of the information it has concerning the world. We begin by simplifying the problem and assuming that each target can only be in one of two states: 1) target is still operational, 2) target is out of commission. Models which admit of a greater but finite number of target states present no special difficulty. The state of the world is then a vector of n components, one for each initial target. The entry is zero for all components which correspond to targets which have been effectively destroyed, and is one for those which are still operational. A bombing action can thus result in any one of the 2^n distinct Boolean vectors. These range from 00 --- 0 to 11 --- 1 and correspond, respectively, to the completely successful mission and to the completely unsuccessful mission. The message or observation vector is a vector of, at most, n components which can also assume the values zero and one.

The command and control system for the bomb-damage-assessment function can be imperfect on one or both of two distinct counts. One, the message or observation vector may be incomplete, i. e., the message vector has fewer components than the state vector. The other, the message vector may have an error, i. e., there may be a non-zero probability that, for at least some of the elements, the message and the corresponding state-of-the-world element do not coincide.

The i-th target is characterized by a threat value t_i and a value v_i. The quantity t_i is negative and represents the risk incurred in not killing the i^{th} target. The quantity v_i is positive and represents the utility of destroying the i-th target.

The weapons used by the organization are assumed to be all identical with cost C and a probability of kill (1-q). Thus, q is the probability of failure.

The acts open to the organization are various allocations of the available weapons on the n potential targets. Thus, the act vector is a vector with n components, the i-th component, a_i, being equal to the number (integer) of weapons allocated against the i-th target.

With the above assumption, the utility which we can associate with an action state-of-the-world pair takes the form:

$$u(A, X) = \sum_{i=1}^{n} [(1 - q^{a_i}) x_i v_i + q^{a_i} x_i t_i - Ca_i] \qquad (I\ 1.1\text{-}1)$$

1.2 Operating Constraints

The sets of acts which are admissible in this organization are those vectors such that:

$$\begin{aligned} I' A &\leq R \\ A &= [A] \geq 0 \end{aligned} \qquad (I\ 1.2\text{-}1)$$

where I is a column vector of n ones, and [A] is the vector of nearest integers. The first constraint expresses the fact that there are R weapons available; thus an admissible act should not utilize more than R weapons. The second expresses the fact that allocations are constituted by integer numbers of weapons for each target and that these numbers are non-negative. The set of A's so constrained is an integral simplex of "Size" R in n+ 1 dimensions. In n dimensional space it is the convex set whose vertices are the origin and the columns (rows) of the indentity matrix multiplied by R.

1.3 The Computation of $\hat{V}(\varkappa)$ for a Generic Bomb-Damage-Assessment System

Using equations (3. 3-1) and (I 1. 1-1), one obtains for $\hat{V}(\varkappa)$ the following expression:

$$\hat{V}(\varkappa) = \sum_{i=1}^{M} \max_{A} \{ \sum_{j=1}^{N} P(X^j) P(Y^i|X^j) \sum_{k=1}^{n} [1-q^{a_k}) x_k^j v_k +$$

$$q^{a_k} x_k^j t_k - Ca_k] \} =$$

$$= \sum_{i=1}^{M} \max_{A} \left[\sum_{k=1}^{n} (1-q^{a_k}) V_k^{Y^i} + q^{a_k} T_k^{Y^i} - Ca_k \right] \qquad \text{(I 1. 3-1)}$$

where

$$V_k^{Y^i} = \sum_{j=1}^{N} P(Y^i|X^j)\, P(X^j)\, x_k^j\, v_k$$

$$T_k^{Y^i} = \sum_{j=1}^{N} P(Y^i|X^j)\, P(X^j)\, x_k^j\, t_k$$

$$C^{Y^i} = \sum_{j=1}^{N} P(Y^i|X^j)\, P(X^j)\, C = C\, P(Y^i)$$

Consequently, to compute the value of $\hat{V}(\chi)$ for our type of organization, one must solve the following type of nonlinear mathematical programming problem:

$$\max_{A} \left\{ \sum_{k=1}^{n} \left[(1-q^{a_k}) V_k^{Y^i} + q^{a_k} T_k^{Y^i} \; Ca_k^{Y^i} \right] \right\}$$

Subject to

$$\sum_{k=1}^{n} a_k \leq R \qquad (a_k = \text{positive integer}) \qquad \text{(I 1. 3-2)}$$

Actually, one has to solve a problem of the type (I 1. 3-2) for each of the distinct messages Y^i. For the sake of simplicity in discussing the solution of (I 1. 3-2), we will replace $V_k^{Y^i}$, $T_k^{Y^i}$, and C^{Y^i} with V_k, T_k, and C, respectively, since Y^i is fixed for each problem (I 1. 3-2). It can be shown (Appendix II) that the Kuhn-Tucker conditions [2] are necessary and sufficient for a problem of the type (I 1. 3-2). The conditions are:

$$g^o + F^{o'} u^o = -w^o$$

For some $u^o \geq 0$ [I]

$$w^o \geq 0$$

g^o being the gradient of the function g computed at the optimal point (act) a^o, i.e.,

$$g^o = [\Delta g(a)]_{a=a^o}$$

F^o is the matrix whose rows are the gradients of the individual constraint surfaces of (I 1.3-2) computed at a^o, i.e.,

$$F^o = \left| \begin{matrix} \Delta f_1 (a)' \\ \Delta f_2 (a)' \\ \vdots \\ \Delta f_m (a)' \end{matrix} \right|_{a=a^o}$$

In our case, the e-th component of g^o is:

$$\frac{\delta}{\delta a_e} \left| \sum_{k=1}^{n} \{(1-q^{a_k}) V_k + q^{a_k} T_k - Ca_k\} \right|_{a=a^o} =$$

$$(T_e - V_e) \log q \, q^{a_e^o} - C$$

Also, since there is a single constraining surface, F^o has a single row whose e-th component is:

$$F_e^o = \left. \frac{\delta F}{\delta a_e} \right|_{a=a^o} = -1$$

Thus [I] in our case is:

$$\left| \begin{matrix} (T_1 - V_1) \log q \, q^{a1} - C \\ (T_2 - V_2) \log q \, q^{a2} - C \\ \vdots \\ (T_n - V_n) \log q \, q^{a_n} - C \end{matrix} \right| + u_1^o \left| \begin{matrix} -1 \\ -1 \\ \vdots \\ -1 \end{matrix} \right| = - \left| \begin{matrix} w_1^o \\ w_2^o \\ \vdots \\ w_n^o \end{matrix} \right|$$

where $u_1^o \geq 0$ and $w_i^o \geq 0$. Furthermore, since the optimal point will be such that $f_1(a^o) = 0$, then $u_1^o > 0$.

Let us assume to begin with that for the optimal act $a_k > 0$ for all k's. Then all the w_i^o are null (Ref. 2) and we obtain the condition,

$$\begin{vmatrix} (T_1 - V_1) \log q \, q^{a_1} - C \\ (T_2 - V_2) \log q \, q^{a_2} - C \\ \cdot \\ \cdot \\ \cdot \\ (T_n - V_n) \log q \, q^{a_n} - C \end{vmatrix} + u_1^o \begin{vmatrix} -1 \\ -1 \\ \cdot \\ \cdot \\ \cdot \\ -1 \end{vmatrix} = 0$$

This is equivalent to saying that all the components of the first vector are equal to each other. This yields n - 1 equations,

$$(T_1 - V_1) \, q^{a_1} = (T_2 - V_2) \, q^{a_2}$$

$$\vdots$$

$$(T_1 - V_1) \, q^{a_1} = (T_n - V_n) \, q^{a_n}$$

The n^{th} equation needed to determine the Lagrangian multiplier is supplied by the constraint, i. e.,

$$\sum_{k=1}^{n} a_k = R$$

A little algebra reduces the above to,

$$\left\{ \begin{array}{l} a_1 = \log_q \dfrac{T_2 - V_2}{T_1 - V_1} + a_2 \\ \vdots \\ a_1 = \log_q \dfrac{T_n - V_n}{T_1 - V_1} + a_n \\ \displaystyle\sum_{k=1}^{n} a_k = R \end{array} \right.$$

which has as a solution,

$$a_i = \frac{1}{n} \sum_{\substack{k=1 \\ k \neq i}}^{n} \log_q \left(\frac{T_k - V_k}{T_i - V_i} \right) + \frac{R}{n} .$$

But, since $\log_q 1 = 0$ one can set

$$a_i = \frac{1}{n} \sum_{k=1}^{n} \log_q \left(\frac{T_k - V_k}{T_i - V_i} \right) + \frac{R}{n}$$

$$a_i = \frac{1}{n} \sum_{k=1}^{n} \log_q \left(\frac{V_k - T_k}{V_i - T_i} \right) + \frac{R}{n}$$

or,

$$a_i = \frac{R}{n} + \log_q \frac{\sqrt[n]{\prod_{k=1}^{n} (V_k - T_k)}}{V_i - T_i} .$$

That is,

$$a_i = \frac{R}{n} + \log_q \frac{G(n)}{V_i - T_i}$$

where $G(n)$ is the geometric mean of the quantities $(V_k - T_k)$, $k = 1, 2, \ldots n$. If on the other hand some of the a_k's should be null, then the corresponding w_k^o are going to be non zero (Ref. 2). Assume that $(n - \lambda)$ a_k's are going to be null, then the K - T condition would read,

$$\begin{vmatrix} (T_1 - V_1) \log q \, q^{a_1} - C \\ (T_2 - V_2) \log q \, q^{a_2} - C \\ \vdots \\ (T_\lambda - V_\lambda) \log q \, q^{a_\lambda} - C \\ (T_{\lambda+1} - V_{\lambda+1}) \log q - C \\ \vdots \\ (T_n - V_n) \log q - C \end{vmatrix} + u_1^o \begin{vmatrix} -1 \\ -1 \\ \vdots \\ -1 \\ -1 \\ \vdots \\ -1 \end{vmatrix} = \begin{vmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ -w_{\lambda+1}^o \\ \vdots \\ -w_n^o \end{vmatrix}$$

The first λ equations together with

$$\sum_{k=1}^{n} a_k = \sum_{k=1}^{\lambda} a_k = R$$

yield

$$a_i = \frac{R}{\lambda} + \log_q \frac{G(\lambda)}{V_i - T_i} \qquad \text{(I 1. 3-3)}$$

where i = 1, 2, . . . λ and

$$G(\lambda) = \sqrt[\lambda]{\prod_{k=1}^{\lambda} (V_k - T_k)}$$

The remaining equations require

$$(T_i - V_i) \log q - C - u_1^o = -w_1^o$$

$$i = \lambda + 1, \lambda + 2, \ldots, n$$

Since $w_i > 0$ in the above equation,

$$(V_i - T_i) \log \frac{1}{q} - C - u_1^o < 0$$

That is:

$$(V_i - T_i) < \frac{u_1^o + C}{\log \frac{1}{q}}$$

$$i = \lambda + 1, \lambda + 2, \ldots, n$$

The first λ equations, on the other hand, yield:

$$(T_j - V_j) \log q \, q^{a_j} - C - u_1^o = 0$$

i.e.,

$$\frac{C + u_1^o}{\log \frac{1}{q}} = (V_j - T_j)\, q^{a_j} = (V_j - T_j)\, q^{\frac{R}{\lambda}}\, q^{\log_q \frac{G(\lambda)}{(V_j - T_j)}} = q^{\frac{R}{\lambda}}\, G(\lambda)$$

Thus, if the i-th target should be allocated no weapons, its equivalent value $(V_i - T_i)$ should be such that

$$(V_i - T_i) < q^{\frac{R}{\lambda}}\, G(\lambda)$$

We are thus in the position to solve any program of the type (I 1.3-2). One would start by computing the quantity

$$q^{\frac{R}{\lambda}}\, G(\lambda)$$

with λ equal to the number of maximum value targets and determining if there are any i's such that,

$$(V_i - T_i) > q^{\frac{R}{\lambda}}\, G(\lambda)$$

If one finds such i's, one would add them to the set of λ targets and compute

$$q^{\frac{R}{\lambda}}\, G(\lambda)$$

computed over the so obtained set of targets. This procedure is iterated until one fails to add any further targets, then utilizing equation (I 1.3-3) one can determine the optimal allocation of weapons $\{a_i\}$ and thus eventually the information system effectiveness.

2. SAMPLE EVALUATIONS

2.1 The Perfect Information System

In this section we will actually compute the effectiveness of the perfect information system for the organization described in Section 1. The perfect information system is represented by a matrix, P, which is the identity matrix.

$$p(X^j|Y^i) = \delta_{ij}$$

where:

$$\delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$ is the Kroenecker delta.

This simply means that the messages and states of the world are in a one-to-one correspondence, i.e., the state of the world is completely and accurately known.

This particular information system is considered for two reasons: a) it establishes an upper bound for the effectiveness of all possible information systems, and b) it poses a simple evaluation problem.

2.1.1 The Computation of $V(\varkappa_o)$

We begin by computing $V(\varkappa_o)$ since the effectiveness of an information system $E(\varkappa)$ is given by $V(\varkappa) - V(\varkappa_o)$.

$\varkappa_o$ is the system characterized by the matrix $Q = \{p(Y^i|X^j)\}$ such that $p(Y^i|X^j) = \frac{1}{N+1}$. Thus, for it we have:

$$V_k^{Y^i} = \sum_{j=0}^{N} \frac{1}{N+1} \; p(X^j) \; x_k^j \; v_k = \frac{E(x_k) \, v_k}{N+1}$$

where $E(x_k)$ is the expected value of x_k.

If the prior distribution is uniform, $E(x_k) = \frac{1}{2}$, since the k-th component of the state vector is equal to one in exactly half of the N+ 1 distinct states and zero for the other half, since $N + 1 = 2^n$ is even. Similarly,

$$T_k^{Y^i} = \frac{E(x_k)\, t_k}{N+1}$$

$$C^{Y^i} = \frac{C}{N+1}$$

Then using equation (1.3-1) one obtains:

$$V(\varkappa_o) = \sum_{i=0}^{M} \max_{A} \left[\sum_{k=1}^{n} (1 - q^{a_k}) \frac{E(x_k)\, v_k}{N+1} + q^{a_k} \frac{E(x_k)\, t_k}{N+1} - \frac{C}{N+1} a_k \right]$$

$$= \frac{M+1}{N+1} \max_{A} \left[\sum_{k=1}^{n} (1 - q^{a_k}) E(x_k)\, v_k + q^{a_k} E(x_k)\, t_k - C a_k \right] \qquad (\text{I } 2.1.1\text{-}1)$$

Assuming that all v_k and t_k are equal, that the number M + 1 of messages and the number N + 1 of states are equal, and that the prior distribution is uniform, one obtains

$$V(\varkappa) = \max_{A} \left[\sum_{k=1}^{n} (1 - q^{a_k}) \frac{v}{2} + q^{a_k} \frac{t}{2} - C a_k \right] \qquad (\text{I } 2.1.1\text{-}2)$$

With the usual constraints

$$I \cdot A \leq R$$

$$[A] \geq 0$$

It can be seen that the solution of this mathematical program is given by:

$$a_k = \left\{ \begin{array}{l} \dfrac{\log - (2\delta)}{\log q} \\[2ex] \dfrac{R}{n} \end{array} \right\} \quad \text{Use the smaller of the two.}$$

where

$$\delta = \frac{C}{\log q\, (v-t)}$$

In the first case one has

$$V(\varkappa_o) = n \left[\frac{v}{2} + C \frac{1 - \log(-2\delta)}{\log q} \right] \qquad (\text{I } 2.1.1\text{-}3)$$

In the second case one has

$$V(\varkappa_o) = n \left[(1 - q^{\frac{R}{n}}) \frac{v}{2} + q^{\frac{R}{n}} \frac{t}{2} \right] - CR \qquad (\text{I } 2.1.1\text{-}4)$$

2.1.2 The Computation of $V(\varkappa_p)$

The system $\varkappa_p$ is the system characterized by the matrix

$$P = [p(Y^i|X^j)] = \delta_{ij}$$

and thus by the matrix

$$Q = \{p(Y^i|X^j)\} = \delta_{ij}$$

where δ_{ij} is the Kroenecker delta. Then for this case one has that:

$$V_k^{Y^i} = \sum_{j=0}^{N} \delta_{ij}\, p(X^j)\, x_k^j\, v_k = p(X^i)\, x_k^i\, v_k$$

$$T_k^{Y^i} = \sum_{j=0}^{N} \delta_{ij}\, p(X^j)\, x_k^j\, v_k = p(X^i)\, x_k^i\, t_k$$

$$C^{Y^i} = \sum_{j=0}^{N} \delta_{ij}\, p(X^j)\, C = p(X^i)\, C$$

Using equation (I 1.3-1) one obtains:

$$V(\varkappa_p) = \sum_{i=0}^{M} \max_{A} \left[\sum_{k=1}^{n} (1-q^{a_k})\, p(X^i)\, x_k^i\, v_k + q^{a_k}\, p(X^i)\, x_k^i\, t_k - C\, p(X^i) a_k \right]$$

$$= \sum_{i=0}^{M=N} p(X^i) \max_{A} \left[\sum_{k=1}^{n} (1 - q^{a_k})\, x_k^i\, v_k + q^{a_k}\, x_k^i\, t_k - C a_k \right] \qquad (\text{I } 2.1.2\text{-}1)$$

In the event that all v_k's and t_k's are equal, this reduces to:

$$V(\chi_p) = \sum_{i=0}^{N} p(X^i) \max_{A} \left[\sum_{k=1}^{n} \{(1-q^{a_k}) v + q^{a_k} t\} x_k^i - Ca_k \right] \qquad (\text{I } 2.1.2\text{-}2)$$

The solution of the problem

$$\max_{A} \left[\sum_{k=1}^{n} \{(1-q^{a_k}) v + q^{a_k} t\} x_k^i - Ca_k \right]$$

$$I \cdot A \leqq R \qquad (\text{I } 2.1.2\text{-}3)$$

$$[A] \geqq 0$$

will depend on one characteristic of the state vector, X^i; namely, the number, λ, of its non-zero components. (This is true because of the assumed homogeneity of the targets.)

If Θ is the set of non-null components of X^i, the unconstrained maximum of the functional is given by:

$$a_k = \begin{cases} \dfrac{\log - \delta}{\log q} & x_k \in \theta \\ 0 & x_k \notin \theta \end{cases}$$

where: $\delta = \dfrac{C}{v - t} \quad \dfrac{1}{\log q}$. The a_k are all positive.

The largest λ for which the unconstrained maximum is the solution of the problem is the largest integer Λ for which

$$\sum_{k=1}^{n} a_k \leqq R \quad \text{or} \quad \Lambda \frac{\log - \delta}{\log q} \leqq R, \text{ i.e.,}$$

$$\Lambda \leqq R \frac{\log q}{\log(-\delta)}$$

Thus, for all the states with $0 \leqq \lambda \leqq \Lambda$

$$\max_{A} \left[\sum_{k=1}^{n} \{(1 - q^{a_k}) v + q^{a_k} t\} x_k^i - Ca_k \right] =$$

$$\lambda(1 - q^{\frac{\log - \delta}{\log q}}) v + \lambda q^{\frac{\log - \delta}{\log q}} t - \lambda C \frac{\log - \delta}{\log q} =$$

$$\lambda \left[v + C\left(\frac{1 - \log(-\delta)}{\log q}\right)\right] \qquad (I2.1.2\text{-}4)$$

If $\lambda > \Lambda$ the maximum is constrained, and it must be obtained by using the K-T condition [I]. If $\bar{\theta}$ indicates the set of null components of X^i, we can begin to use the K-T conditions to determine whether any weapons have to be allocated against the corresponding targets. First, one has to compute

$$q^{\frac{R}{\lambda}} G(\lambda)$$

where $k \in \theta$

λ being, as indicated above, the number of members of θ. In this case,

$$G(\lambda) = v - t$$

Then, if

$$(V_j - T_j) < q^{\frac{R}{\lambda}} (v - t) \qquad j \in \bar{\theta}$$

one would allocate no weapons against the targets corresponding to $\bar{\theta}$. But if $j \in \bar{\theta}$, $(V_j - T_j) = 0$, and since $q^{\frac{R}{\lambda}} (v - t)$ is positive in all cases, we can conclude that the constrained optimum will require allocating no weapons to those targets which do not exist, which of course is an expected result. The remaining K-T conditions for $k \in \theta$ require that the projection of the gradient of the functional in the subspace corresponding to the indexing set θ be parallel to the vector of all ones, which, in turn, is equivalent to requiring

$$(v - t) \log q \; q^{a_e} = (v - t) \log q \; q^{a_k} \qquad a_k \in \theta$$

$$\sum_{k \in \theta} a_k = R$$

These are solved by $\quad a_k = a = \frac{R}{\lambda}$

In this case then

$$\max_{A} \left[\sum_{k=1}^{n} \{ (1 - q^{a_k}) v + q^{a_k} t \} x_k^i - C a_k \right] =$$

$$\lambda \left[(1 - q^{\frac{R}{\lambda}}) v + q^{\frac{R}{\lambda}} t - C \right] \qquad (I\ 2.1.2\text{-}5)$$

The expected utility of the perfect information transformation for homogeneous targets can be computed using equations (I 2.1.2-2), (I 2.1.2-4), and (I 2.1.2-5) to be:

$$v(\chi_p) = \sum_{\lambda=0}^{\Lambda} \Pi(\lambda) \; \lambda \; \left[v + C\left(\frac{1 - \log(-\delta)}{\log q}\right) \right] +$$

$$\sum_{\lambda=\Lambda+1}^{n} \Pi(\lambda) \; \lambda \; \left[(1 - q^{\frac{R}{\lambda}}) \; v + q^{\frac{R}{\lambda}} \; t - C \right]$$

$$v(\chi_p) = \left[v + C \; \frac{1 - \log(-\delta)}{\log q} \right] \sum_{\lambda=0}^{\Lambda} \Pi(\lambda) \; \lambda +$$

$$\sum_{\lambda=\Lambda+1}^{n} \Pi(\lambda) \; \lambda \; \left[(1 - q^{\frac{R}{\lambda}}) \; v + q^{\frac{R}{\lambda}} \; t - C \right] \qquad (I\ 2.1.2\text{-}6)$$

2.1.3 Sample Results

At this point, subtracting equation (I 2.1.1-3) or (I 2.1.1-4), whichever applies, from equation (I 2.1.2-6) one obtains

$$\epsilon(\varkappa_p) = V(\varkappa_p) - V(\varkappa_o)$$

Graphs 1 - 4 show $V(\varkappa_p)$ and $V(\varkappa_o)$ as calculated for our organization using the following combinations of parameters listed below.

	#1	#2	#3	#4
No. targets, n	4	4	4	4
Target value, v	5000	5000	5000	5000
Target threat, t	-5000	-5000	-5000	-5000
Weapon cost, C	1000	1000	1	1
No. avail. weap., R	16	4	100	8
Prior distribution	Uniform	Uniform	Uniform	Uniform

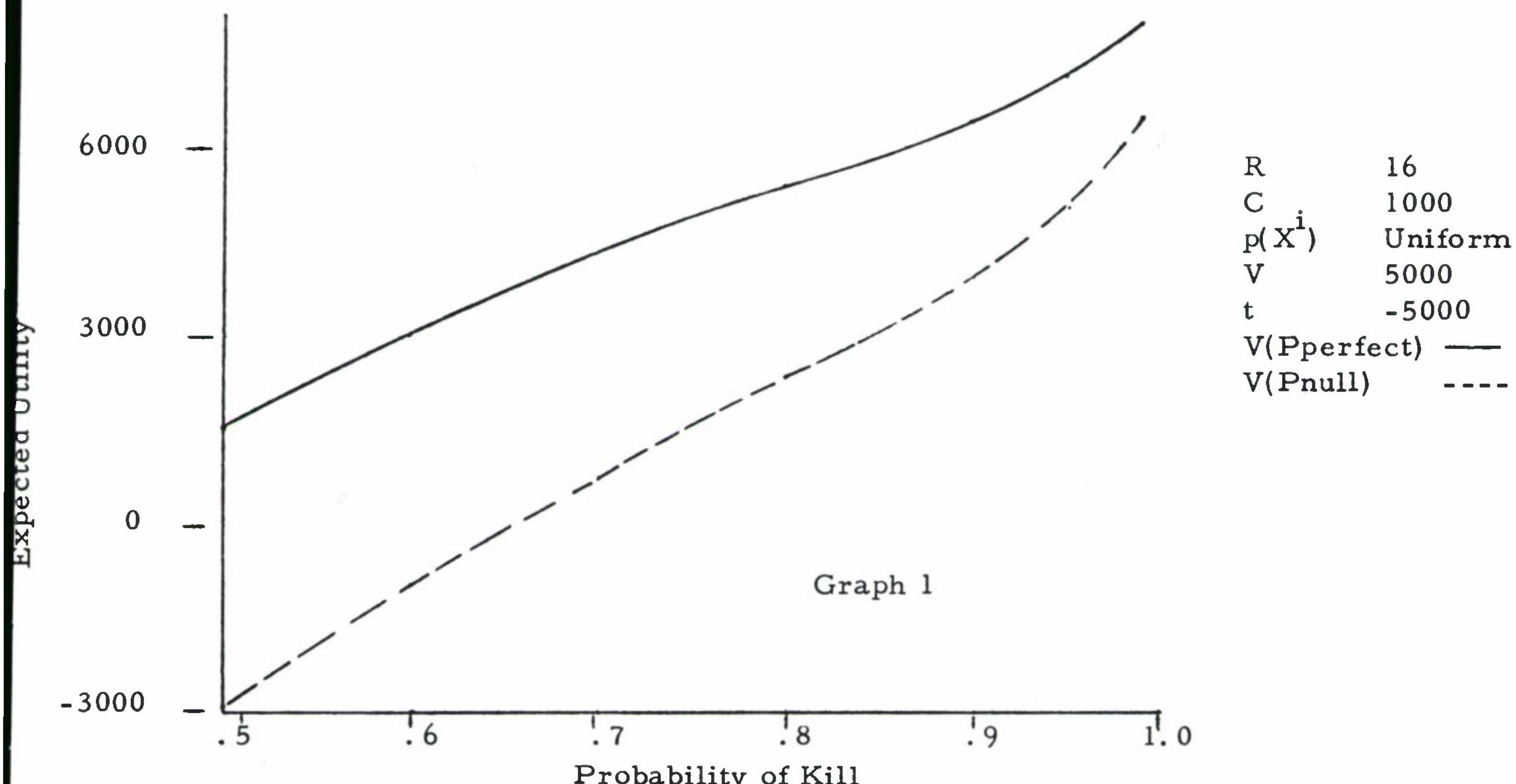
The above values were chosen to create representative cases. As it can be seen in all cases, the organization faces up to 4 targets, all of equal value and threat value. The cost of the weapon is the first variable in the table. Graphs 1 and 2 correspond to the case where the weapon is worth 1/10 of the equivalent value of the target. These graphs, therefore, are more relevant for decision making involving the allocation of weapon systems rather than single weapons. The second group of graphs, on the other hand (numbers 3 and 4), with a weapon cost three orders of magnitude smaller, can be considered relevant for decision making when allocating single weapons. The graphs are also differentiated on the basis of whether resources are abundant or scarce. The abundant and scarce cases are given in graphs 1, 3 and 2, 4, respectively. In all cases, the prior distribution is uniform.

Graph 3 represents decision making involving the use of cheap resources in ample supply. It is clear from this graph that there is essentially no benefit to be gained from adopting a bomb-damage-assessment function.

Graph 4 shows the effect of a limited supply for a cheap resource. As it can be seen the effectiveness of the information system increases with a decrease in the reliability of the weapon employed. The limited supply of weapons makes it impossible to saturate the targets, thus demonstrating the desirability of having bomb-damage information in order to use one's limited weapons most effectively.

The perfect information system is seen to be capable of bringing about a maximum of 80% improvement in the performance of the organization. In graph 1, we see the performance of the organization when its decision making entails the use of expensive resources, such as entire weapon systems, which are in abundant supply. In graph 2, the limited supply case, the improvement of performance is considerable, and one observes that the effectiveness of the perfect information system attains a maximum for weapon-kill probabilities in the range .6 to .7. More representative data could be obtained if the prior state distribution used could be the binomial distribution for the targets previously killed. Such curves could be obtained if the above procedure were computerized.

PERFECT INFORMATION



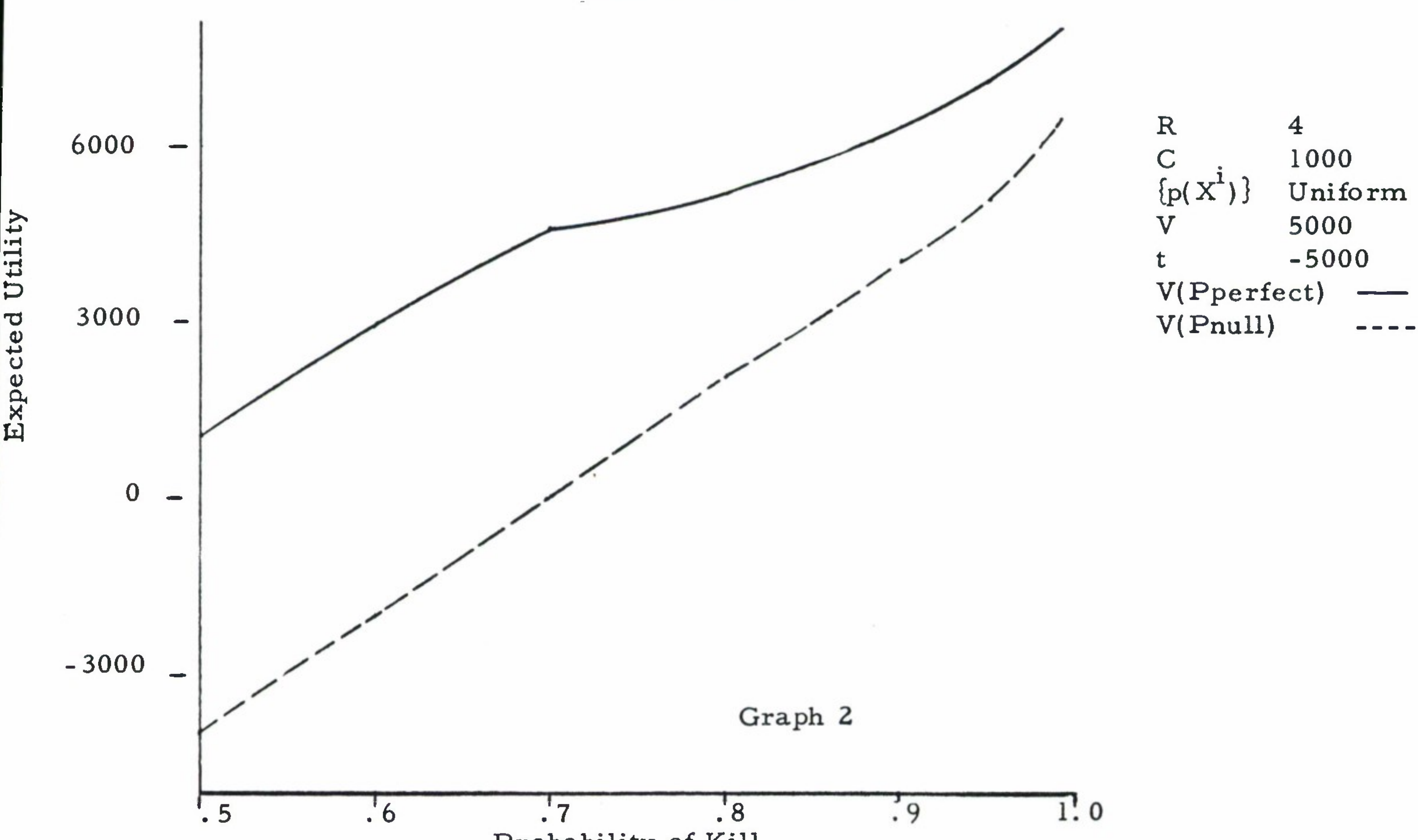
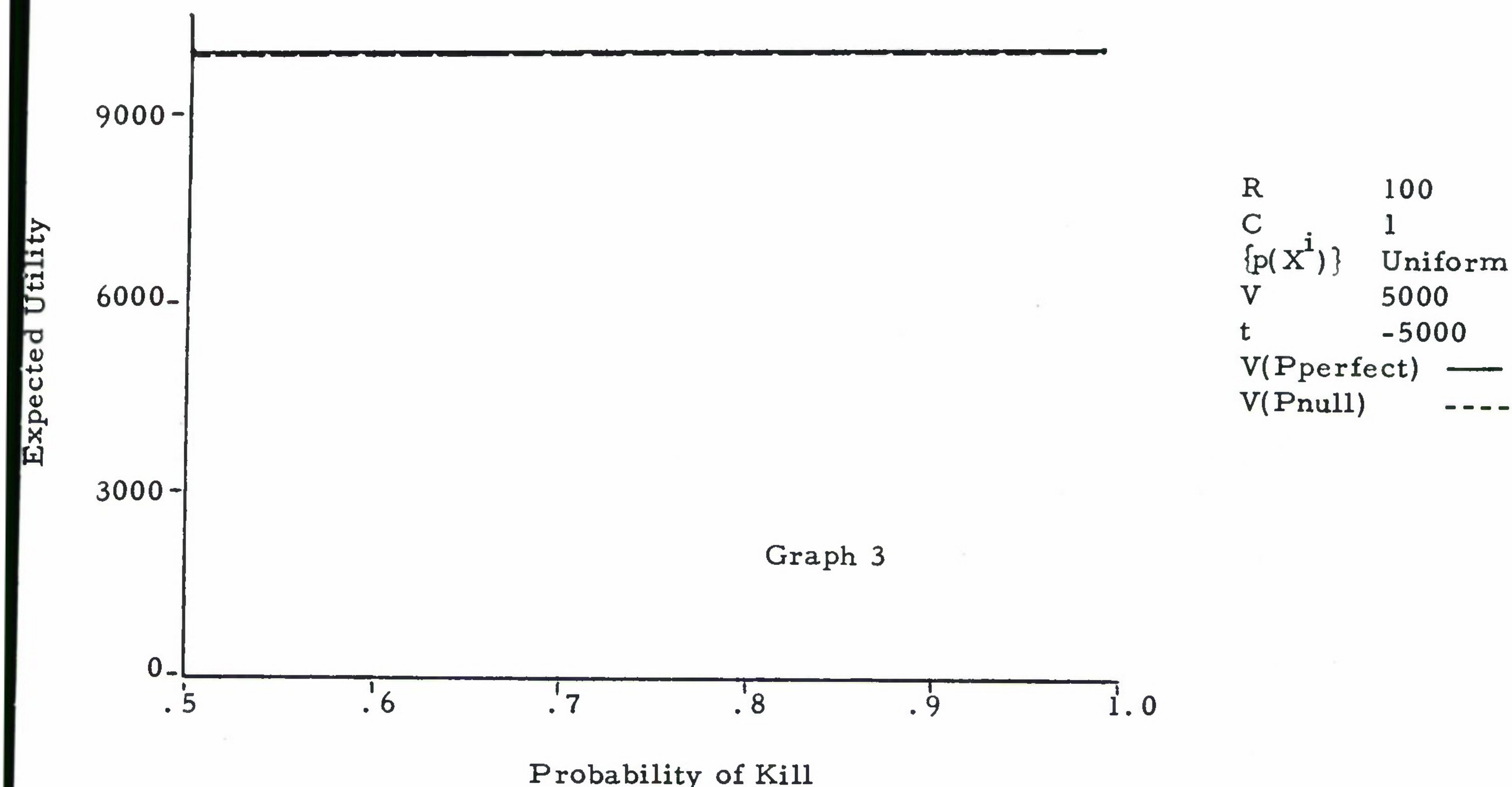


Figure I-1.

PERFECT INFORMATION



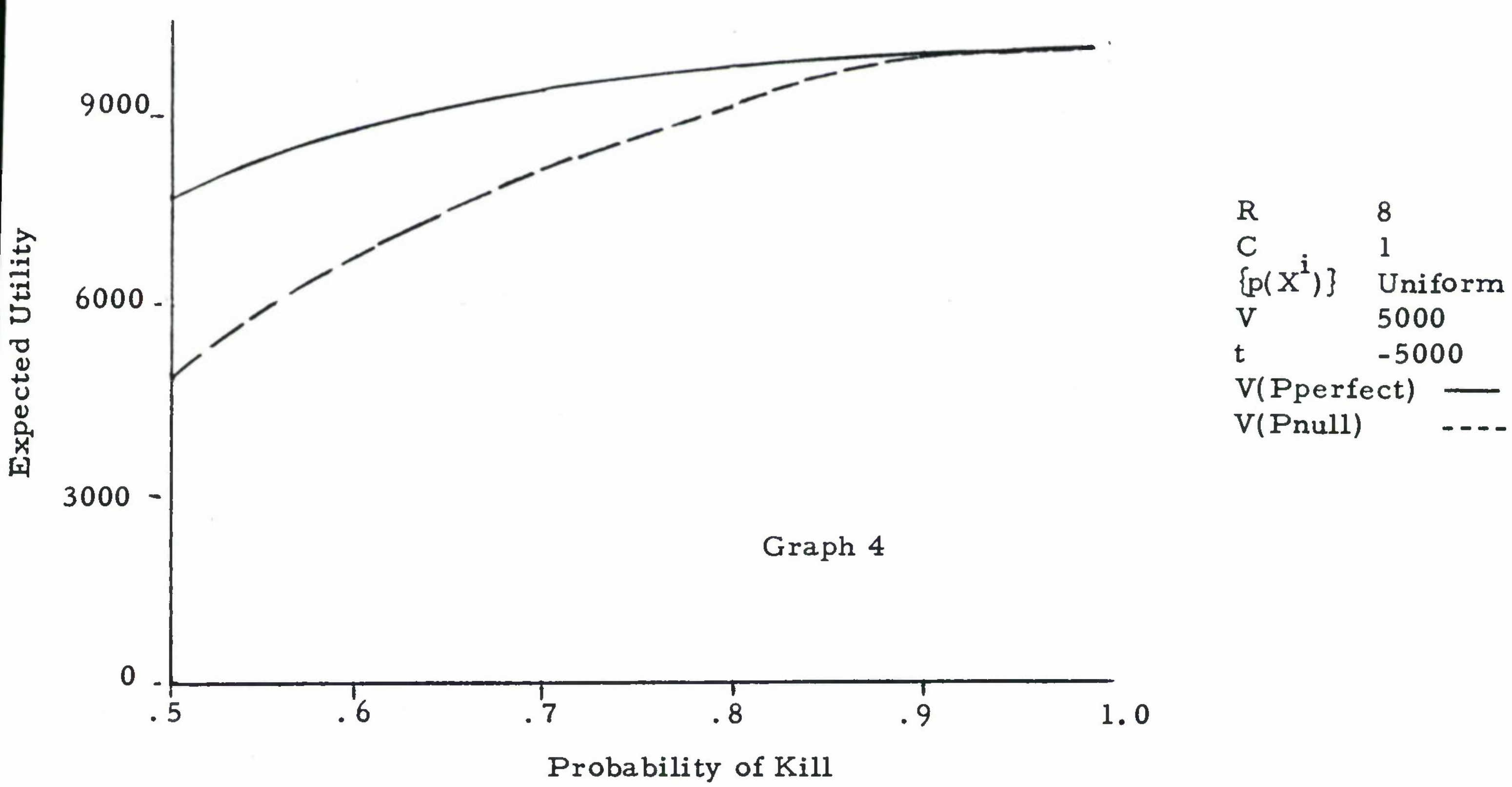


Figure I-2.

APPENDIX II

THE KUHN-TUCKER CONDITIONS

In this appendix some of the key ideas concerning the solution of the problem

$$\begin{cases} \max & g(x) \\ Fx \geq & 0 \\ x \geq & 0 \end{cases} \qquad \text{(II-1)}$$

are reviewed. In (B-1) g(x) and

$$Fx = \begin{vmatrix} f_1(x) \\ f_2(x) \\ \cdot \\ \cdot \\ \cdot \\ f_m(x) \end{vmatrix}$$

are assumed to be differentiable functions in the positive orthant. We will begin by deriving the Kuhn-Tucker conditions as the necessary conditions for solving the mathematical programming problem as given by (II-1).

Fx is mapping defined by the m differentiable functions $f_i(x)$, i = 1, 2, . . . m. Consequently, the matrix F^o

$$\begin{vmatrix} \frac{\delta f_1}{\delta x_1} & \frac{\delta f_1}{\delta x_2} & \cdots & \frac{\delta f_1}{\delta x_n} \\ \frac{\delta f_2}{\delta x_1} & \frac{\delta f_2}{\delta x_2} & \cdots & \frac{\delta f_2}{\delta x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\delta f_m}{\delta x_1} & \frac{\delta f_m}{\delta x_2} & \cdots & \frac{\delta f_m}{\delta x_n} \end{vmatrix}_{x=x^o} = F^o$$

exists (F^o is computed at $x = x^o$), and so does

$$g^o = (\Delta g(x))_{x=x^o}$$

where x^o satisfies the constraints of problem (II-1). Incidentally, F^o can be written also as

$$F^o = \begin{vmatrix} (\Delta f_1(x))' \\ (\Delta f_2(x))' \\ \cdot \\ \cdot \\ (\Delta f_m(x))' \end{vmatrix}$$

where the various gradients are represented as column vectors.

In particular, if we assume that x^o is on the boundary of the set $Fx \geq 0$, $x \geq 0$, some of these conditions will be satisfied at the equality level. Consequently, we may partition the two sets of conditions in four sets as follows:

$$Fx \geq 0 \quad \begin{cases} F_1x \geq 0 & \text{corresponding to} \quad F_1x^o = 0 \\ F_2x \geq 0 & \text{corresponding to} \quad F_2x^o > 0 \end{cases}$$

$$x \geq 0 \quad \begin{cases} I_1x \geq 0 & \text{corresponding to} \quad I_1x^o = 0 \\ I_2x \geq 0 & \text{corresponding to} \quad I_2x^o \geq 0 \end{cases}$$

F_1x is the mapping obtained from Fx by striking out those functions $f_i(x)$ for which

$$f_i(x^o) > 0$$

F_2x is simply the list of the so eliminated functions. I_1 and I_2 are matrices obtained from the identity matrix by collecting the rows corresponding to null components of x^o and non-null components, respectively.

If x^o is a maximum for problem (II-1), then it must be true that

$$g^{o'} dx \leq 0$$

for all feasible dx's. In fact, $g^{o'} dx = (\Delta g)_{x=x^o} \cdot dx$, is dg along dx. The feasible or allowed dx are those which do not cause an exit from the feasibility region. Now, we have that

$$dx = x - x^o$$

and

$$Fx = Fx^o + F^o(x - x^o)$$

since $dx = x - x^o$ is assumed to be infinitesimal. Consequently, for the conditions involving the F_1x part of the mapping, we have

$$F_1x = F_1x^o + F_1^o(x - x^o) = F_1^o(x - x^o)$$

since by definition $F_1x^o = 0$. Finally, since $F_1^o x \geq 0$, it follows that $F_1^o(x = x^o) = F_1^o\, dx \geq 0$. Similarly, the $x \geq 0$ constraint yields for the displacement dx the condition

$$I_1\, dx \geq 0.$$

Thus, we can say that for x^o to be a solution of (II-1) it is necessary that

$$g^{o'}\, dx \leq 0 \qquad \text{for all } dx \qquad \text{(II-2)}$$

$$F_1^o\, dx \geq 0$$

$$I_1\, dx \geq 0$$

At this point we apply the following theorem due to Farkas:

$$b'x \geq 0 \text{ for all } x\text{'s} \ni Ax \geq 0 \text{ only if}$$

$$b = A't \qquad \text{for some } t \geq 0$$

From (II-2) and this theorem, one obtains

$$-g^o = \begin{vmatrix} F_1^o \\ I_1 \end{vmatrix}' \qquad t = \begin{vmatrix} F_1^o \\ I_1 \end{vmatrix}' \begin{vmatrix} u_1^o \\ w_1^o \end{vmatrix}$$

for some $u_1^o \geq 0$, $w_1^o \geq 0$, or

$$-g^o = F_1^{o'} u_1^o + I_1' w_1^o \text{ for some } u_1^o \geq 0,\ w_1^o \geq 0 \qquad \text{(II-3)}$$

and adding appropriate zeros to u_1^o and w_1^o to form u^o and w^o, one can write (B-3) as

$$-g^o = F^{o'} u^o + w^o \text{ for some } u_1^o \geq 0,\ w^o \geq 0 \qquad \text{(II-3')}$$

Thus we can conclude that the condition

$$g^o + F^{o'} u^o = -w^o \qquad w^o \geq 0,\ u^o \geq 0 \qquad \text{(II-4)}$$

where

$$g^o = (\Delta g(x))_{x=x^o}$$

$$F^o = \left| \begin{array}{c} f_1\ (x)' \\ f_2\ (x)' \\ \cdot \\ \cdot \\ f_m(x)' \end{array} \right|_{x=x^o}$$

is a necessary condition for x^o to be a solution of

$$\left\{ \begin{array}{l} \text{Max } g(x) \\ f_1\ (x) \geq 0 \\ f_2\ (x) \geq 0 \\ \quad \cdot \\ \quad \cdot \\ f_m\ (x) \geq 0 \\ X \geq 0 \end{array} \right. \qquad \text{(II-1')}$$

where the f_i (x) and g(x) are required to be differentiable in $X \geq 0$. In our case g(x) is (strictly) concave and the $f_i(x)$ are linear. Let's prove that under these circumstances (II-3') is sufficient also. From Reference [2] sufficient condition for x^o to be a solution for (II-1) is that

(I) $\Phi_x^o \leq 0 \quad \Phi_x^{o'} x^o = 0 \quad x^o \geq 0$ and

(II) $\Phi_u^o \geq 0 \quad \Phi_u^{o'} u^o = 0 \quad u^o \geq 0$ and

(III) $\Phi(x, u^o) \leq \Phi(x^o, u^o) + \Phi_x^{o'}(x - x^o)$

For $\Phi(x, u) = g(x) + u'Fx$

Furthermore, Φ_x^o and Φ_u^o are the gradients of $\Phi(x, u)$ with respect to, respectively, x and u and computed at the (x^o, u^o) point. We will begin by showing that in our case (III) is indeed verified.

The assumption that g(x) is a concave function is equivalent to stating that

$$g(x) \leq g(x^o) + g^{o'}(x - x^o) \tag{II-5}$$

The assumption that the $f_i(x)$'s are linear functions is equivalent to stating that

$$f_i(x) = f_i(x^o) + f_i'(x^o)(x - x^o)$$

and thus that

$$Fx = Fx^o + F^o(x - x^o) \tag{II-6}$$

If (II-6) is premultiplied by $u^{o'}$ and added to (II-5), one obtains

$$g(x) + u^{o'} Fx \leq g(x^o) + u^{o'} Fx^o + (g^{o'} + u^{o'} F^o)(x - x^o) \tag{II-7}$$

Now it can be easily verified that

$$\Phi_x^o = g^o + F^{o'} u^o$$

Thus, (II-7) is

$$\Phi(x, u^o) \leq \Phi(x^o, u^o) + \Phi_x^{o'}(x - x^o)$$

which is condition III. Since Reference 2 shows (I) and (II) can be derived from (II-3'), this establishes the sufficiency of (II-3'). Thus, to state (II-3') is in our case equivalent to stating that x^o is a solution of the mathematical programming problem (II-1).

In addition, it is well worth noting the following properties. The derivation of (II-3') in Reference [2] makes it clear that the vector u^o has zeros in correspondence to those $f_i(x)$'s which enter in the constraints which are not satisfied at the equality level; i. e., for each

$$f_i(x^o) > 0$$

there is a corresponding zero in u^o. Similarly, for each

$$x_i^o > 0$$

(x_i^o is the i^{th} component of the vector x^o) there is a null component in the w^o vector.

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* Dunlap and Associates, Inc., Darien, Conn.	2a. REPORT SECURITY CLASSIFICATION N/A 2b. GROUP

3. REPORT TITLE

MATHEMATICAL PROGRAMMING TECHNIQUES FOR INFORMATION SYSTEM DESIGN

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

5. AUTHOR(S) *(Last name, first name, initial)*

Gagliardi, U. O. Ying, C.C. Holt, L. G.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
July 64	80	5

8a. CONTRACT OR GRANT NO. AF19(628)2830 b. PROJECT NO. 2804 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) DRD-64-112 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* ESD-TDR-64-530

10. AVAILABILITY/LIMITATION NOTICES

OTS release authorized.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Electronic Systems Division, AFSC USAF

13. ABSTRACT

The problem of establishing the effectiveness of an information system is considered. An effectiveness measure suggested by recent development in statistical decision theory is presented. Sample evaluations of system designs are used to illustrate here the adoption of such a measure which allows selecting the parameters of the system in a manner consistent with the user preference.

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Command Control Systems Statistical Analysis Computers Operations Research Communication Systems						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through
_______________________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through
_______________________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through
_______________________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS), (S), (C), or (U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.